

Refocusing NASA's vision

Ballooning costs for NASA's next space telescope are putting other worthwhile projects at risk — and carry lessons for future mission planning.

It is now clear that when NASA first requested funds for a Next Generation Space Telescope a decade ago, the project's advertised price of under \$1 billion was little more than a fiction. But it was a fiction that the space agency, the Congress and many in the astronomy community wanted to believe.

The consequences of this self-deception are now coming home to roost (see page 140). The James Webb Space Telescope (JWST), as the project is now known, is by far the largest item in NASA's science programme. Its serious cost overruns, together with the space agency's transfer of money from the science budget to support future astronaut missions, are causing the indefinite postponement of other exciting projects. And while smaller missions can be cancelled if they break their budgets — Dawn, NASA's mission to the asteroid

"If the telescope's current price had been on the sticker when it was first offered for sale, the project might never have got off the drawing board."

belt, suffered just this fate only last week — a flagship mission such as the JWST is all but immune to such a drastic disciplinary measure.

This chain of events has cast an unfortunate light on a means of project selection that has often been exalted as a model for consensus building between competing research projects: the Decadal Survey for astronomy, conducted by the US National Academy of Sciences. Back in 2000, the Decadal Survey backed seven major projects and a dozen or so smaller ones, and suggested that \$4.6 billion would pay for the lot of them. The \$1-billion project that was to become the JWST was top of the list. NASA's best estimate for the cost of this instrument alone is now \$4.5 billion.

A cynic might suppose that the telescope's budgetary trajectory has been cleverly crafted; in the beginning it looked small enough to be affordable, and once under way it quickly became too big to be cancelled. If the telescope's current price had been on the sticker when it was first offered for sale, the project might never have got off the drawing board.

All inclusive

The same could be said, in fact, of the much-admired Hubble Space Telescope, which the JWST will in some regards replace. It is by no means certain that Congress would have given the go-ahead for Hubble had it known that its final lifetime cost, including servicing missions on the space shuttle, would reach about \$12 billion at today's prices.

That figure for Hubble, although suitably astronomical, does at least have the benefit of being fully inclusive: it contains overheads, operations, inflation and everything else. One of the reasons that the JWST now looks so expensive is that it is the first large NASA project to be scrutinized under the agency's new policy of declaring costs in this way.

Painful as such declarations may be, they will ultimately serve the agency better than the prevarications of the past. But good accounting will not in itself ensure that funding for major projects flows in the orderly manner that is needed if they are to stay on the right budgetary track. At the moment, it is impossible to predict whether NASA's science budget will remain stable or be cut further as the costs of a new human exploration programme and the legacy of current obligations add up. The previously envisaged growth in the budget certainly seems unlikely to materialize.

In such circumstances, flagship missions as ambitious as the JWST — missions to study planets round other stars, or return samples from the surface of Mars, for example — will be few and far between, if any fly at all. International cooperation will doubtless be explored as a way forward, and this should involve genuine collaboration at the planning stage and cost-sharing from the outset, rather than the secondary roles that Europe and Canada play in the JWST. But such collaborations require real political commitment at the very highest level and, as the International Space Station amply demonstrates, they suffer from many drawbacks of their own.

A question of balance

Even if costs are shared internationally, there is a limit to how much funding space-based astronomy can expect to receive compared with other scientific disciplines. Astronomers may say it is not their fault that their tools are more expensive than those of other scientists; beryllium mirrors unfolding in deep space inevitably cost a great deal more than butterfly nets. But while it is right that every dimension of human curiosity should have some call on the public purse, it is not reasonable to expect the guardians of that purse to choose between the costly and the cheap with complete equanimity. Even with the current cuts, NASA's science receives more than \$5 billion a year, and few outside space science regard that as insufficient.

The astronomy community must therefore learn to live within its means, starting with its next Decadal Survey. Princeton physicist Joseph Taylor, who co-chaired the last one, told Congress last week that the once carefully crafted plan is becoming unbalanced: "The current NASA budget," he said, "is tilted to an unhealthy extent toward the large missions." Given the choice, most in the community would probably prefer any funds freed up by the ending of old projects to go on small, focused missions and research grants, rather than on the development of future super-missions. As a result, the next survey will need to consider a more balanced portfolio of less heroic ambition. And it will have to look at it with an accountant's squint, as well as an astronomer's far-seeing gaze.

"The astronomy community must therefore learn to live within its means, starting with its next Decadal Survey."



A warm welcome

Scientists should embrace a move by evangelicals to join the debate on climate change.

President George W. Bush likes to talk about his faith at every opportunity; then he jokes about the poor grades he received in science class. Bush, a methodist, experienced his shift towards religion late in life. But now he may find the ghosts of his long-lost science education rising up to meet him on Sunday mornings.

Bush identifies with the hazily defined but devout group of Americans known as evangelicals, who also form the bedrock of his electoral support. Loosely speaking, these are people who believe in the authoritative word of the Bible above all else. But now, it seems, the word of science textbooks may be gaining some authority of its own among the evangelicals.

Scientists should welcome the recent move by leading evangelicals to call for action on climate change (see page 136). Their statement provides an unprecedented opportunity for science to make a real impact on a broad segment of US society.

Americans as a whole are deeply devout, a fact that has perplexed environmental groups that operate from an agnostic or atheist perspective. But the disconnect is not just about religion. Talk to people in America's heartland and you will find an underlying distrust of academics from either coast. Any indication that there might be a problem with the climate — and that pick-up trucks might have something to do with it — is likely to be viewed with hostility or

suspicion. This is a culture in which people like to do things their own way, and they resent mandates that they perceive as being handed down by city-based élites.

Evangelicals themselves feel alienated from much of the country; one 2004 poll showed that 48% of evangelicals feel they are looked down on by most of their fellow Americans. Scientists should take the opportunity to reach out to this group.

The United States is unlikely to sign up to any successor agreement to the Kyoto treaty to cut greenhouse-gas emissions without a groundswell of support from its citizens. The evangelical leaders' open approach to the issue is a clear signal that it's okay to be religious, devout even, and still believe that climate change is a major problem.

With a blessing from religious leaders, Americans no longer need to feel divided between their conservative beliefs and their concern for the planet. The evangelicals' call to action draws on a long history of Christian environmental activism, founded on the notion of caring for Earth as a gift from God, to be cherished for future generations. But in this case it confronts one of the most politically charged issues of the day.

The evangelicals' declared interest in the climate-change issue is an important approach that climate scientists should welcome with open arms. The scientists should engage with them at every possible level, starting at their local church. Perhaps, in due course, even President Bush will become part of the discussion. ■

"The evangelical leaders' approach is a clear signal that it's okay to be religious, devout even, and still believe that climate change is a major problem."

Slow train coming

Reform of Ukraine's archaic research system is needed sooner rather than later.

Ukraine suffers from an identity crisis that is inhibiting its scientific, as well as its economic and political, development. The 48 million inhabitants of the former Soviet republic are deeply divided between pro-European and pro-Russian factions. The celebrated 'orange revolution' of November 2004 did less to bridge this divide than is commonly thought.

The nation's research system broadly reflects this wider societal divide. On the one hand, there are many young, well-educated and highly motivated researchers and a network of increasingly independent universities. On the other, there's the National Academy of Sciences of Ukraine, a leviathan of militant senility that retains just enough power to control critical aspects of Ukraine's scientific life.

The academy employs 45,000 permanent staff in a network of largely unproductive research establishments. Given the advanced age of its senior management, time alone will eventually resolve the issue. But that won't happen soon enough for those young Ukrainians currently in search of a productive scientific career.

Integrating the Ukraine into the Framework research programme of the European Union (EU) would allow this generation far greater interaction with its peers abroad. The European Commission

supports the idea, which could also help open the way to future EU membership for Ukraine. But the leadership of the academy, deeply rooted in Soviet traditions, seems to be thwarting such integration through a mixture of contrariness and lack of interest (see page 132).

A high-level EU-Ukrainian steering committee on scientific cooperation, for example, was established on paper four years ago but has yet to actually meet. When it does, the academy's leaders are expected to obstruct collaborative steps that might bring an infusion of foreign influences into the country — including respect for the value of independent peer review.

Ukrainian science has potential in several spheres, including materials sciences, radioastronomy, theoretical physics and agricultural research. The nation badly needs to focus its scarce resources in those areas where its scientists can compete, and dispose of some of its anachronistic scientific heritage. That will require a rigorous external evaluation of the performance of hundreds of the academy's institutes.

The government needs to identify these reforms as a priority and then act with determination to overcome the academy's likely resistance to them. The oligarchy that has controlled Ukrainian science since Soviet times may then lose out. But the nation's economic potential and its prospects for integration into the EU, as well as science itself, can only benefit. ■

"The nation badly needs to focus its scarce resources in those areas where its scientists can compete."

RESEARCH HIGHLIGHTS

POPULATION GENETICS

Celtic mutations

Nature Genet. doi:10.1038/ng1742 (2006)

Symptoms of the motor neuron disease amyotrophic lateral sclerosis (ALS) occur in mice whose gene for the neuroprotective protein VEGF is damaged. But mutations in this gene have never been found in human ALS patients.

A surprising link between ALS and VEGF, which also promotes the growth of new blood vessels during hypoxia, has now been made in a multi-population study. It included more than 1,600 ALS patients, and was run by Dublin-based scientists Matthew Greenway and Orla Hardiman of the Royal College of Surgeons in Ireland and their colleagues.

In 15 individuals — nearly all of Irish or Scottish descent — the study identifies mutations in the gene for angiogenin, a protein that seems to be required for normal VEGF activity in blood vessels.

CONSERVATION

Mix it up

Proc. R. Soc. Lond. B doi:10.1098/rspb.2006.3477 (2006)

Isolated small populations are prone to reduced fitness because they are forced to inbreed — a phenomenon that threatens to become more widespread as natural habitats are destroyed. Conservationists, however, have avoided strategies that link adjacent isolated populations because of the risk of spreading disease.

Researchers led by John Hogg of the Montana Conservation Science Institute in Missoula now provide some evidence to inform the debate. Their data, spanning 25 years, document fitness improvements in a flock of bighorn sheep (*Ovis canadensis*;

Under pressure

Geophys. Res. Lett. **33**, L03312 (2006)

When a lava dome collapsed at Soufrière Hills Volcano in 2003 — the biggest such collapse on record — it triggered a sudden and dramatic increase in pressure in the underlying magma chamber.

During the event, researchers led by Barry Voight of Pennsylvania State University in University Park measured strain around the volcano (pictured), which has been erupting on the Caribbean island of Montserrat since 1995. The magma chamber pressurized within ten minutes of the lava dome's collapse, much faster than would be expected for the normal rate of magma flow.

The scientists suggest that bubbles within the magma may have expanded suddenly, causing the pressure rise.



S. O'MEARA & D. O'MEARA/SPL

pictured below) consisting of a few dozen animals, and isolated since 1922. Since the introduction of outsiders in 1985, fitness measures such as birth weight and male and female reproductive success have improved — gains that, the authors argue, offset the risks.

CHEMISTRY

Metal tools

J. Am. Chem. Soc. **128**, 2540–2541 (2006)

Organometallic chemists can often 'tweak' the reactivity of a transition-metal catalyst by adding ligands that bind to the metal — an approach that researchers have now used to expand the synthetic chemist's tool box.

Barry Trost and his co-workers at Stanford University, California, report that a palladium catalyst with a phosphorus-containing ligand can catalyse the oxidation, by a nitronate, of an ester group adjacent to a carbon-carbon double bond. If the ligand was a single enantiomer — a molecule that is not superimposable on its mirror-image — the product of the reaction was also a single enantiomer. This was true even when the starting material was a mixture of enantiomers. The products of this reaction could be used to synthesize various natural products, such as prostaglandins.

ANATOMY

Unexpected organ

Science doi:10.1126/science.1123497 (2006)

After years of study, the laboratory mouse can still surprise.

Hans-Reimer Rodewald at the University of Ulm, Germany, and his colleagues say they have discovered a new organ — a thymus the size of a pin-head — in the necks of mice. Previously, the mouse was thought to have just one thymus, situated near the heart.

The thymus helps to supply the immune system with T cells, and this unexpected finding raises questions about studies that have used mice with the main thymus removed to investigate, for example, T-cell production in other organs, such as the gut and skin.

MATERIALS SCIENCE

Pretty cool

Science **311**, 1270–1271 (2006)

A material that can be cooled by applying an electric field might help to prevent microchips from overheating, say Alex Mischenko of Cambridge University, UK, and his colleagues.

The researchers find that applying 25 volts



J. HOGG

across a thin film of lead zirconate-titanate can cause its temperature to drop by up to 12 °C at around 220 °C. This effect could be exploited to create an electrically driven heat pump. Lead zirconate-titanate is a well known piezoelectric material, meaning that an applied voltage creates a mechanical stress. But precisely why this can also lead to cooling is unclear. Previous observations of the 'electrocaloric effect' have been more than 100 times weaker — too small to be of much practical use.

OPTICS

Loopy fibres

J. Phys. B **39**, 1011-1016 (2006)

Physicists in France have shown that the optical Sagnac effect, which is used in rocket navigation systems, also exists at the quantum level.

The effect is revealed when a laser beam is split and passed in opposite directions around a loop of rotating optical fibre. The interference pattern that is formed when the beams recombine depends on the fibre's angular velocity, so the system can be used as a rotation detector. Guillaume Bertocchi and his colleagues at the University of Nice Sophia-Antipolis detected a similar link between angular velocity and the interference pattern that built up as they sent single photons round the loop. The effect, however, may be too small to be of practical use.

CANCER

Unleashing the guardian

Nature Chem. Biol. doi:10.1038/nchembio774 (2006)

A drug that boosts the activity of our cells' anti-cancer guardian, the p53 protein, has an extra trigger that helps it

to target only tumour cells.

p53 usually forces cells to commit suicide if they start to become cancerous. But in half of all tumours, the protein is ineffectual. Nutlin-3, a small molecule, is one of a class of anti-cancer drugs that kills such tumour cells by activating p53.

Nutlin-3 works by preventing p53 from being gagged by another protein, known as MDM2. But, René Bernards and Roderick Beijersbergen at the Netherlands Cancer Institute in Amsterdam and their colleagues have found that the effectiveness of nutlin-3 also depends on the activity of the gene *53BP1*. This gene forms part of a DNA-damage signalling system, which is often turned on in cancerous cells.

ACOUSTICS

To be is to didgeridoo

J. Acoust. Soc. Am. **119**, 1194-1204; 1205-1213 (2006)

Researchers in Australia have delved into the country's native heritage to reveal the physics of the didgeridoo, or *yidaki*. A team led by Joe Wolfe of the University of New South Wales in Sydney has used a range of methods to study the techniques used to play the instrument. The work builds on a study published in *Nature* last year (A. Tarnopolsky *et al.* **436**, 39; 2005).

The didgeridoo, although it is a simple resonating tube, can produce a remarkable array of different sound effects, chiefly through the player opening and closing their glottis, or voicebox. Wolfe and his team also measured the acoustic effects of moving the lips and of using circular breathing. They publish an accompanying theoretical analysis.

Sound files are available at www.phys.unsw.edu.au/~jw/didgeridu.html.

RNA INTERFERENCE

Brain protectors

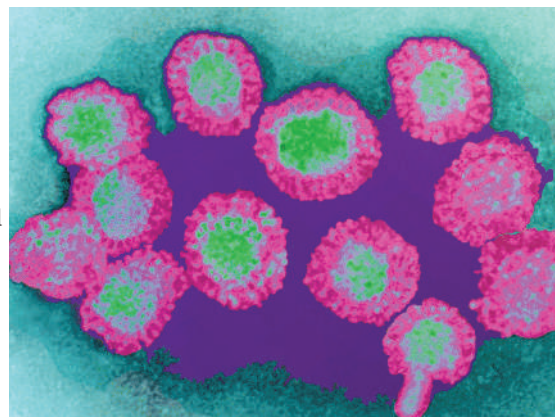
PLoS Med. **3**, e96 (2006)

A scrap of RNA may halt the brain inflammation caused by the Japanese encephalitis virus and the related West Nile virus (pictured below), for which there are no effective treatments.

Researchers led by Premalata Shankar and Manjunath Swamy, both of Harvard Medical School in Boston, Massachusetts, tried combating the viruses by the technique of RNA interference.

They injected tiny fragments of RNA directly into the brains of mice. The RNA, delivered into cells by a lentivirus or lipid, blocked the activity of a key protein present in the coat of both viruses so that the viruses could not replicate.

The technique might work in humans if scientists can figure out a less invasive way of getting the RNA into the brain.



JOURNAL CLUB

Victor Smetacek
Alfred Wegener Institute for
Polar and Marine Research,
Bremerhaven, Germany

A bio-oceanographer learns a lesson from elephants.

I once watched a lone bull elephant pushing over young trees planted in an extensive stretch of grassland bordering the Himalayas. The gunshot-like cracks of snapping trunks accompanied his stately passage through the plantation.

Many years later, I learned that the wanton destruction was actually adaptive behaviour. This demolishing of trees maintains grassland — their preferred food.

But trees can outgrow the rate of destruction when rainfall is above 650 mm per year, as shown by a recent assessment of savanna and closed-canopy forest habitats in Africa (M. Sankaran *et al.* *Nature* **438**, 846-849, 2005).

I am struck by the parallels with marine plankton, my field of study. Diatom blooms are the equivalent of closed-canopy forests in the

sea: they overgrow their zooplankton grazers when the supply of iron is adequate. Blooms, fed by iron leaching from the continents, are prominent along ocean margins, but drop off beyond the shelf break. Adding iron to these deeper, nutrient-rich (but iron-poor) ocean waters is like watering the savanna: diatom blooms take over.

Diatom chains subsequently sink out in flakes, like falling leaves, into the depths of the ocean. This creates a carbon deficit in the surface waters, which is

compensated by draw-down of carbon dioxide from the atmosphere. So fertilizing the ocean with iron could sequester some anthropogenic carbon dioxide, an idea being tested in experiments.

But on land the opposite seems to hold. Some suggest that when trees take over grasslands, the amount of carbon sequestered will go down. So, perhaps letting elephants restore their former habitats would contribute more to carbon dioxide sequestration than would planting forests.

NEWS

Ukraine scientists grow impatient for change

Ukraine's 'orange revolution' — a national protest against corruption that overthrew the first results of the country's 2004 election — raised hopes for political and societal change. But more than a year on, scientists are increasingly frustrated by the slow pace of reform of the country's Soviet-style research system, which they believe is being hampered by Ukraine's aged and anti-European scientific establishment.

The nation, which has a population of 48 million and is Europe's second-largest country in terms of area, has a long tradition in science and hosts

an extensive network of academic institutes and research facilities. But, as it did elsewhere in Eastern Europe, science declined dramatically after the collapse of communism in 1991, forcing thousands of researchers to leave the country.

When Viktor Yushchenko came into power in January 2005, it was hoped that the pro-West president would encourage a fundamental reform of the science system. But critics say that the promised switch to less a authoritarian system has hardly begun.

The focal point of criticism is the National Academy of Sciences of Ukraine (NASU), which runs 174 institutes and employs around

28,000 researchers. The powerful academy, a relic of the Soviet science complex, dominates Ukrainian science. The average age of the academicians is about 71; the president, Boris Paton, an expert in electric welding and the son of the former president, is 85.

The bulk of the academy's activities relate to mechanics, material sciences and physics — euphemisms, according to critics, for former

"Nothing will change in Ukrainian science as long as the current system exists."

military-oriented engineering institutes. And productivity is low. According to the Thomson Scientific (ISI) statistics, academy scientists publish around 1,500 papers a year — roughly one-third of the output of Britain's University of Manchester alone.

But critics say the academy is not interested in carrying out an independent review of its scientific performance. There are also claims of widespread corruption. For example, an attempt to create closer ties between Ukraine and western European institutions by linking Ukraine to GÉANT, the high-speed European data communication network, was allegedly hindered by academy members demanding bribes. Another complaint is that the academy leaders, fearing competition and loss of influence, are blocking attempts to facilitate Ukraine's participation in research pro-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Russian premier Vladimir Putin (left) salutes academy president Boris Paton, aged 85.

grammes funded by the European Union (EU), by deliberately holding back information and generally failing to cooperate with EU authorities.

"The Academy is not interested in any reform whatsoever," says Aleksei Boyarski, a theoretical physicist at CERN, the European lab for particle physics in Geneva, Switzerland.

E. LUKATSKY/AP/GETTY IMAGES

Evidence for bubble fusion called into question

ONLINE SPECIAL REPORT

Fresh questions surround the claims that bubble fusion has been achieved, according to an investigation by *Nature*.

Reports by Rusi Taleyarkhan that he had achieved table-top fusion in collapsing bubbles caused a storm when they were published in 2002 (R. P. Taleyarkhan *et al. Science* 295, 1868–1873; 2002).

Taleyarkhan, a nuclear engineer now based at Purdue University in West Lafayette, Indiana, used sound waves to cause the formation and collapse of bubbles within a liquid. The conditions inside the collapsing bubbles are theoretically extreme

enough to allow nuclear fusion to take place. Taleyarkhan claims to have achieved this — an effect that, if real, could one day provide an almost limitless source of energy.

Four years later, Taleyarkhan's work retains an almost magical ability to grab the headlines, most recently in January, when his latest results (R. P. Taleyarkhan *et al. Phys. Rev. Lett.* 96, 034301; 2006) were promoted in a press release by the American Physical Society. Millions of dollars are being spent trying to repeat the work, including \$800,000 from the US Department of Defense.

But corroboration remains elusive. Now, an investigation into

the circumstances surrounding Taleyarkhan's experiments is throwing up serious questions about the validity of the work.

Interviews with researchers who have worked closely with Taleyarkhan at Purdue reveal concerns about his actions since he arrived there full-time in 2004. The steps he has taken, they say, include claiming he obtained positive results from equipment on which they had seen only negative data, and removing the equipment from their lab altogether.

And physicist Brian Naranjo of the University of California, Los Angeles, has completed an analysis that he

plans to post later this week on arXiv. It suggests that the spectrum reported in Taleyarkhan's latest paper as proof of nuclear fusion came instead from the radioactive decay of a standard lab material.

Taleyarkhan has declined to comment on events at Purdue, or on Naranjo's analysis, and he vigorously affirms that his results are valid and the effect is real. But the overall message from people close to this work is that there is little hope this particular approach will yield a viable fusion energy source. ■

Eugenie Samuel Reich
For the full investigation please see
▶ www.nature.com/news/bubblefusion



"Nothing will change in Ukrainian science as long as this system exists."

Ukrainian scientists are eligible for EU research money thanks to a 2002 association agreement with the European Commission's framework programme for research. But so far, only seven out of thousands of EU-funded projects include Ukrainian participants, says Vadym Yashenkov, deputy director of Ukraine's National Information Point for EU research.

According to Yashenkov, this is partly because of the general weakness of Ukrainian science and industry, and the complicated application procedures that put off many scientists.

But participation is also hindered because the academy fails to provide and disseminate relevant documents and information, says Oleh Napov, a science attaché at the Ukrainian mission to the EU in Brussels, Belgium. For example, Napov has submitted a proposal for scientific reform to the Ukrainian research ministry. He says that when he asked the academy to outline its scientific priorities, he received only a list of the names and titles of all current academicians, and a letter stating that the academicians themselves were the academy's priorities.

"Maybe they have not asked us in a proper way," counters Yaroslav Yatskiv, director of the Main Astronomical Observatory in Kiev, and a member of the academy's presidium. The president, Paton, had not responded to queries when *Nature* went to press.

Yatskiv says he is aware that corruption is a widespread problem within the academy. "It is true, unfortunately, that funding is not based on scientific merit," he says. But he adds that efforts to evaluate and possibly transform the academy are being considered.

Brain drain

Yatskiv has recently proposed the creation of a National Science Foundation that, like its US counterpart, would fund research on the sole basis of excellence judged by peer review. But Paton last year told a presidium meeting that the future role of the academy should be similar to that of the Siberian branch of the Russian Academy of Sciences, another relic of the Soviet science complex. "I don't think this is a good idea," says Yatskiv.

Resistance to the academy's backward-looking plans is growing, both inside and outside Ukraine. A 13-strong group of Ukrainian scientists, led by Boyarski, has suggested to the country's science ministry a detailed concept of domestic reform, including rigorous evaluation of all academy institutes, the creation of an international institute of advanced study in Kiev and of a number of centres of excellence supported by the EU.

"Things back home really need to improve substantially," says Alexej Verkhatsky, a Ukrainian-born neurophysiologist at the University of Manchester and a member of Boyarski's group. "If they don't, our best young people will soon have left for good. A considerable number of Ukrainian scientists working abroad (myself included) would come back if things were reorganized."

"We have the same potential, scientifically and politically, as Poland or Hungary to become a genuine part of Europe," adds Oleg Krishtal, deputy director of the academy's Bogomoletz institute of physiology in Kiev. "What we need is proper political stimulus. Clearly, the academy cannot repair itself as long as the old guard is keeping all the key positions."

Christian Patermann, director for biotechnology, agriculture and food at the European commission's directorate general for research in Brussels, led an EU delegation to Ukraine last month. He says that the country's scientific potential in areas such as materials sciences, energy, space and organic farming is impressive and deserves European support. Patermann is optimistic that the academy will not ultimately stand in the way of reform. "The Czech Republic, Hungary and the Baltic countries have all managed to reform their academies of science; sooner or later this will also happen in Ukraine." ■

See also page 128.

Quirin Schiermeier

ON THE RECORD

"Mars is hard, and Mars is unpredictable. Mars doesn't treat you very well."

NASA's Jim Graf crosses his fingers for the Mars Reconnaissance Orbiter spacecraft, due to arrive at the red planet on 10 March.

"Libertines, both male and female, have always been around in math and physics."

Science writer Jennifer Ouellette ponders whether physicists are really having more sex these days.

Sources: *Washington Post*, *Seed*

SCORECARD



Science and kung fu

Movie star Jackie Chan has just secured himself a very different type of billing. The Australian National University is naming a new science centre after the action hero as a vote of thanks for his recent donation.



Waterfowl

Researchers in Canada have come up with a way to stop wading birds landing on oil fields in Alberta. As the birds fly in, radar picks them up and triggers a deterrent noise.



Perfume

Japanese scientists have found an unusual source of vanillin, the aromatic component of vanilla used as a fragrance. The sweet smell can, apparently, be extracted from cow dung.

NUMBER CRUNCH

A study by the US Centers for Disease Control and Prevention suggests that moving to the United States, especially if you are hispanic, might be bad for your health.

16% of Hispanic immigrants living in the United States for five years or less are obese.

13% of the same group have high blood pressure.

22% of Hispanic immigrants living in the United States for more than five years are obese.

20% of the same group have high blood pressure.



Squaring up: arguments about artists' use of lenses have focused on the tapestry in this 1543 painting by Lorenzo Lotto.

Tempers blaze over artistic integrity

Sparks flew when a scientist and a renowned artist joined forces to argue that great Renaissance painters used lenses to create their masterpieces. Five years on, an even fiercer controversy burns, as one of them accuses his chief critic of misconduct.

The travail began when painter David Hockney and optics scientist Charles Falco at the University of Arizona, Tucson, published a book and several articles laying out a remarkable theory. They argued that major European masters from the Renaissance to the nineteenth century had used lenses or mirrors to project images on to their canvases, to capture certain details of expression or perspective more accurately (see *Nature* **412**, 860; 2001).

In one line of evidence, Falco and Hockney used lenses to recreate particular scenes in paintings. They saw distortions that matched those present in the original pictures, for example in the octagonal pattern of a tapestry in Lorenzo Lotto's *Husband and Wife* (see above).

Their work spawned an international circuit of conferences and widespread public interest. David Stork, optics expert and chief scientist at Ricoh Innovations in Menlo Park, emerged as the most vociferous and frequent critic, arguing that the masters didn't "cheat".

After a particularly acrimonious meeting in January 2005, Falco says he sat down to dissect the arguments in Stork's articles and lectures. He discovered what he claims are fabrications and manipulations that constitute misconduct. Stork "systematically used erroneous data", Falco has written in letters of complaint.

He adds that Stork's articles "are at significant variance with accepted standards of scientific publication".

For instance, Stork has analysed the *Husband and Wife* scene too, and claims that Lotto did not use a lens. But part of his argument uses a photo of a carpet with an octagonal pattern, and Falco claims that Stork selected a photo of one with a distorted weave. When this warping is corrected, says Falco, the results support his own theory. Falco also alleges that Stork fabricated two of five data points relating to a Georges de la Tour painting of Christ.

Stork told *Nature* he was shocked by the charges. "I categorically deny any inappropriate inactivity," he says. "This is insane." Christopher Tyler, director of the Smith-Kettlewell Brain Imaging Center in San Francisco and co-author of Stork's Lotto analysis describes his role in the study as "light", but says that he stands by the work. Several other experts contacted by *Nature* at Stork's request were unable to provide further information about Falco's allegations.

In May, Falco sent details of his analysis and a letter requesting an investigation to the publishers of nine of Stork's articles, which mostly appeared in the proceedings of conferences where Stork had spoken. He also sent details to Stanford University in California. Stork has widely represented himself as associated with Stanford, using a university e-mail address and web page for his criticisms. But although Stork

has taught occasional courses at Stanford, he is not now affiliated with the university.

After eight months in which he got no significant response from Stanford or any of the publishers, Falco says he decided to publicly disclose the allegations. The decision was partly spurred by a recent review that Stork wrote for *Nature* in which, Falco alleges, his theory was mischaracterized again (see *Nature* **438**, 916; 2005 and *Nature* **439**, 392; 2006). "He attacks Hockney's and my professional competence in press releases and in talks that he solicits for himself based on his erroneous publications," claims Falco.

Editors of the International Society for Optical Engineering, the Institute of Electrical and Electronics Engineers, the Optical Society of America, and the Society for Imaging Science and Technology are among those who have published

Stork's articles. They say they have no plans to investigate the matter, a stance that Falco describes as "disappointing".

But in November, Stanford removed Stork's web page and e-mail from its server. And Stork acknowledges that in December the university's research dean, Arthur Bienenstock, asked him not to represent himself as being at the university. Bienenstock has made no official comment about whether the university will investigate the claims, but Falco said last week that he is encouraged his allegations are being taken seriously.

Rex Dalton

Can cats spread avian flu?

Felines are fast becoming a new focus for fears over avian flu, as cats infected with the deadly H5N1 strain are reported in Austria, Germany, Thailand and Indonesia. So could they spread the virus? The World Health Organization (WHO) has played down the danger based on current knowledge, but experts warn that the science is moving rapidly.

The Austrian authorities announced on 6 March that three domestic cats had tested positive for H5N1 in the southern town of Graz, the scene of a recent outbreak in birds. That followed detection of the virus in a dead cat on the northern island of Rügen, Germany, on 28 February, and news that 8 of 111 apparently healthy cats tested close to bird flu outbreaks in central Thailand carried antibodies to the virus (see *Nature* **439**, 773; 2006).

In a statement last week, the WHO maintained a careful but reassuring tone: "There is no present evidence that domestic cats play a role in the transmission cycle of H5N1 viruses. To date, no human case has been linked to exposure to a diseased cat."

That is all true, for now. In February 2004, the WHO reported the first outbreak in domestic cats. H5N1 was found in two of three cats tested from a household of 15 cats (of which 14 died) in Nakornpathom, Indonesia. At the time the WHO argued that cats are not naturally susceptible to flu, and that even if infected they would not shed large quantities of virus.

But with bird flu it may be different. Later in 2004, Albert Osterhaus's team from Erasmus University in Rotterdam showed experimen-

tally that domestic cats do die from H5N1 and do transmit it to other cats (T. Kuiken *et al. Science* **306**, 241; 2004). And in January this year, the virus was found not only in sputum but also in faeces of experimentally infected cats, suggesting that infected animals may shed the virus extensively (G. F. Rimmelzwaan *et al. Am. J. Pathol.* **168**, 176–183; 2006).

It is unclear how these findings relate to cats in their natural environment. But in next month's issue of *Emerging Infectious Diseases*, Thai researchers describe a cat that died of H5N1 after eating a pigeon carcass. It showed similar pathology to cats experimentally infected with the virus.

Meanwhile, Andrew Jeremijenko, head of influenza surveillance at the US Naval Medical Research Unit 2 in Jakarta, Indonesia, detected H5N1 in a kitten he found near a poultry outbreak in Cipedang, West Java, and tested out of curiosity on 22 January. The virus from the kitten is closely related to recent H5N1 strains isolated from humans in Indonesia: it shares genetic changes found in human strains that are not present in samples from birds.

But scientists may just be learning what is already common knowledge among Indonesian villagers. Peter Roeder, a consultant for the UN Food and Agriculture Organization, says locals have an onomatopoeic name for bird flu "that sounds like 'plop', the sound of a chicken hitting the ground when it falls out of a tree. They also have a name for the cat form of avian flu — 'aaargh plop' — because cats make a screaming noise before they fall out of the tree." ■

Declan Butler

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Feline fears: cats are testing positive for H5N1, but the significance of this remains unclear.

SPECIAL REPORT

Church joins crusade over climate change

Evangelical leaders have called on the United States to step up its efforts to control greenhouse-gas emissions. But can they force action where others have failed, asks **Amanda Haag**.

Fire and brimstone are coming to the aid of US science, as evangelical scientists and their allies in the religious community embark on a battle against climate change.

"The time has come...for destroying those who destroy the Earth," says Calvin DeWitt, a professor of environmental studies at the University of Wisconsin, Madison, quoting from the Scriptures. The Bible teaches stewardship of the planet, he says, which is partly why 86 prominent US evangelical Christians last month signed the 'Evangelical Climate Initiative' calling for mandatory limits on greenhouse-gas emissions.

The movement began in 2000, when 50 evangelical scientists — including DeWitt — signed a statement calling for policy-makers to take steps towards reducing the threat of climate change. It is a rare move in the United States, where environmentalists and the religious community often find themselves in opposite camps. Climate activists hope the

initiative will have the political clout to help sway President George W. Bush's administration towards mandatory emissions cuts. Bush has not signed up to the international Kyoto Protocol on regulating greenhouse gases. Instead, he is promoting clean-energy technologies through agreements such as the six-nation Asia-Pacific partnership. Yet many of Bush's core supporters are religious conservatives.

Evangelicals are a powerful social force in the United States, with the National Association of Evangelicals (NAE) boasting 30 million members. 'Evangelical' is an umbrella term encompassing more than 50 denominations whose members typically believe in Jesus Christ and that the Bible is the authoritative word of God. The NAE has not officially endorsed the climate initiative, but many of the organization's leaders believe it represents a growing consensus that climate change is a matter for concern.

Climate researchers are watching the



movement with optimism. Jim White, a University of Colorado geochemist who studies ice sheets in Greenland, says that it will almost certainly accelerate public support for action on climate change. "To have a group that has historically fought the notion come

The man who preaches science

John Houghton, former co-chair of the scientific working group for the Intergovernmental Panel on Climate Change and a lay preacher in his home church, has long worked to marry scientific understanding of the environment with Christian principles.

Many agree that Houghton's talk at a 2002 conference in Oxford, UK, which drew 70 leading climate scientists, policy-makers and Christian leaders, marked a turning point in mobilizing evangelicals to take a lead in climate change. "I had a conversion experience and just did a 180 on the subject," says Richard Cizik, vice-president of government affairs for the NAE. Jim Ball,

executive director of the Evangelical Environmental Network, calls Houghton and other evangelical scientists "our early prophets on this problem".

The US leaders who attended the Oxford conference told Houghton they were impressed with the humility with which he and other scientists spoke about their science, emphasizing the uncertainties.

In 1997, Houghton and other Christian scientists set up the John Ray Initiative, an educational charity to promote environmental stewardship in accordance with Christian principles. John Ray was an eminent seventeenth-century naturalist and theologian.



John Houghton: works to build bridges with industry over the environment.

Houghton travels extensively, lecturing on environmental stewardship and working to build bridges with business and industry. He has just returned from India, where as a trustee of the Shell

Foundation — a charity that encourages sustainable energy in the developing world — he was involved in a programme to build and market less-polluting, higher-efficiency cooking stoves for villages.

Calling on the biblical principle that "to whom much is given, much will be required", Houghton teaches that people in the wealthier parts of the world have derived huge benefits from fossil fuels and have a responsibility to the developing world. "One of our God-given tasks as humans is to care for creation," he says. "That, I think, is a strong imperative right at the beginning of the Bible."

A.H.



Climate-induced changes in sea level are seen by many evangelicals as having a moral dimension.

around — I think that does impact on the public's thinking," he says.

And it is this public support that some believe could influence conservative legislators. Eileen Claussen, president of the Pew Center on Global Climate Change, based in Arlington, Virginia, says that the lack of significant public demand for action on climate has hampered acceptance by members of Congress. "So I think the evangelical initiative is welcomed by all," she adds.

One reason many are hopeful about change is that certain key evangelicals — such as Ted Haggard, president of the NAE, and Richard Cizik, the group's vice-president of government affairs — speak regularly with the White House. "That's very significant," says DeWitt. "I think the president really wants evangelicals to see him as evangelical." If Bush does not warm to the idea of mandatory emissions, DeWitt says, he could lose some of his key support.

Backers of the initiative are now distributing the statement to Congress, hoping to educate legislators on their views. "They may be able to hear the message about climate change from

us where they couldn't necessarily hear it and really listen to it from others," says Reverend Jim Ball, executive director of the Evangelical Environmental Network, which is championing the initiative.

A positive development would be to bring together two bipartisan groups of senators — one led by John McCain (Republican, Arizona) and Joe Lieberman (Democrat, Connecticut), and the other by Pete Domenici (Republican, New Mexico) and Jeff Bingaman (Democrat, New Mexico) — who have separately put forward climate-change legislation. Domenici and Bingaman have scheduled an April congressional climate conference to discuss ways of mandating emissions cuts.

The evangelicals say they realize they won't change minds overnight. "We're under no illusions that our statement, or its circulating, is going to break open the log jam," says Ball. In the meantime, the group is airing advertisements on major television networks. They next aim to build relationships within the business community, in part by planning for a November meeting between business leaders



SURPRISE ORGAN DISCOVERED IN MICE
Mice are shown to have two thymus organs, not just one.
www.nature.com/news

and evangelicals to help show that climate change can be tackled in ways that will not harm the economy.

John Houghton, a leader in the Christian environmental movement (see "The man who preaches science"), says the task is particularly hard in the United States. He lectures frequently to international audiences and says that, outside the United States, he rarely encounters resistance to the validity of climate-change science. But leaders of the initiative feel the science is now solid enough to convince even the unbelievers. "If there was not such an overwhelming scientific consensus, we probably wouldn't be able to get traction on this issue in our community," says Ball.

Biblical imperative

But there is still plenty of ground to cover. The NAE's Cizik opted not to sign the statement, although he is a convert to accepting climate change as a reality and helped persuade many supporters of the initiative. Cizik had originally signed the document, but 22 evangelical leaders asked him not to be seen as taking a stand for the NAE, which historically acts only in cases of consensus on an issue. Cizik retracted his name, feeling that he could make a stronger case as a facilitator than an advocate.

For him, the ramifications are greater than politics alone. "I believe the very reputation of the gospel is at stake," he says. He likens climate change to the civil-rights movement of the 1960s, in which evangelicals did not act aggressively.

A union between evangelicals and scientists was only a matter of time, says DeWitt, who has written at length on "evangelical environmentalism". Raised in the Christian Reformed Church, he grew up believing that investigation of the natural world goes hand in hand with biblical theology. Not until he went to college did he become aware of the divide between the two communities. "We've built this illusion that we can talk about ourselves on the one hand and the environment on the other hand," says DeWitt.

For many evangelicals, the flashpoint was the growing realization that climate change could wreak its worst effects on the poorest countries, in the form of heat waves, floods and tropical diseases. Sea-level rise could immerse low-lying regions, and agricultural productivity could be sharply reduced in areas such as sub-Saharan Africa.

More than ever, evangelicals are viewing their call to respond as a biblical and moral imperative. "It's a bigger question now," says DeWitt. "Do you really answer to the creator of Heaven and Earth?" ■

P. JOHNSON/CORBIS

Japan anticipates green light for nuclear plants

Japan is edging closer to launching its controversial nuclear-fuel recycling system; the first local approval for a power plant using recycled mixed-oxide fuel (MOX) looks almost certain.

The Japanese government decided in 1997 to promote the use of MOX — a mix of plutonium and uranium produced by recycling spent nuclear fuel. But progress in adopting the fuel has been slow, due to a fatal accident at a nuclear power reactor, scandals involving falsified inspection data, and strong resistance from local people over safety concerns.

But the first green light for using MOX, expected to come from the conservative government of Saga prefecture on the southern island of Kyushu, could speed the introduction of such plants around the nation. It could also help the government justify the nearly ¥2 trillion (US\$17.2 billion) it is spending on a fuel-recycling plant in Rokkasho village in the northern prefecture of Aomori. This is scheduled to go into operation next year.

After a prefectural assembly signalled its approval last month, local officials are likely to endorse the plan to use MOX at one of four nuclear-power reactors in the town of Genkai before the assembly closes on 23 March. They will then make a formal agreement with a local power company, which is expected to be confirmed within the next few months. It is hoped that the plant will be operating by 2010, initially using MOX bought from France or Britain.

The agreement will be a huge relief for the government, as recycling nuclear fuel is at the core of its long-term energy policy. Japan has few natural energy resources. And because it currently gets around a third of its energy from nuclear power, it also needs a way to get rid of the nuclear waste produced. The government wants to replace uranium with MOX at 16 to 18 of Japan's 52 nuclear power plants by 2010.

Critics argue that the plutonium in recycled fuel would release more radiation than conventional fuel in the event of an accident, and would be a target for terrorists. "Security measures at Japanese nuclear power plants

are not nearly as great as they should be," says Edwin Lyman, a physicist at the Union of Concerned Scientists in Cambridge, Massachusetts.

France, Germany, Switzerland and Belgium all run MOX plants, with much of the fuel produced at

the Britain's plants in Sellafield. All but France have decided to stop their operations. But many believe Saga officials are being encouraged partly by the US government's announcement on 6 February of a \$250-million plan to resume research of reprocessing spent nuclear fuel, a major reversal of a 1970s policy banning the use of such fuel. ■

Ichiko Fuyuno

"Security measures at Japanese nuclear power plants are not nearly as great as they should be."

Plant biologist named as Australia's science adviser

The Australian government has appointed plant researcher Jim Peacock as the nation's new chief scientist. He replaces Robin Batterham, whose reign was marred by allegations of conflicts of interest, as he worked part-time for mining giant Rio Tinto while advising the government on energy and climate issues (see *Nature* 435, 398; 2005).

The position is part-time, something that Peacock had previously criticized in his role as president of the Australian Academy of Science (AAS). "Our position is still that it's such an important job that it should be a full-time job," says Sue Serjeantson, executive secretary of the AAS. Peacock will spend the rest of his time conducting research at the Commonwealth Scientific and Industrial Research Organisation.

Environmental groups are concerned about Peacock's enthusiasm for genetically modified crops and for encouraging debate about nuclear energy.

Kurt Lambeck, a geophysicist at the Australian National University in Canberra, will replace Peacock as the AAS president when his term ends in May.

EMBO helps life scientists set up labs in Europe

The Heidelberg-based European Molecular Biology Organization (EMBO) has launched a funding scheme to help talented life scientists set up laboratories in EMBO member states whose science bases are relatively undeveloped. Croatia, the Czech Republic, Poland, Portugal and Turkey

have so far signed up to the programme.

EMBO will select winners of what will be known as the EMBO Strategic Development Installation Grants, and enrol them in its Young Investigator Programme, a sort of finishing school for elite biologists.

Awardees will receive €50,000 (US\$60,000) per year. EMBO is asking host institutes to commit to employing scientists more permanently when the grant period — probably three to five years — has ended.

Japanese researcher finds synthetic route to Tamiflu

A University of Tokyo researcher says he has made a synthetic version of Tamiflu, thought to be the most effective drug against avian influenza.

The Swiss company Roche, which makes Tamiflu using a plant extract, has been unable to meet the huge demand from governments that are stockpiling the drug as avian influenza spreads.

"This will make it possible to have stable production," says Masakatsu Shibasaki, the biochemist who devised the new production method. It uses a readily accessible chemical — 1,4-cyclohexadiene — instead of the plant extract. Shibasaki says the product is exactly the same as Tamiflu, and Tokyo University is beginning to negotiate with Roche over a possible collaboration. A representative from Chugai Pharmaceutical, the company's Japanese subsidiary, says that Roche is aware of Shibasaki's announcement, but that the company cannot release details of any negotiations.

Tokyo University applied for a patent on 23 February. Shibasaki will present his results at the Pharmaceutical Society of Japan's annual meeting on 28–30 March.



Looking up: \$2.2 million could help researchers find and protect the ivory-billed woodpecker.

Friends of lost woodpecker hope for cash windfall

US officials are seeking \$2.2 million to help conserve the 'rediscovered' ivory-billed woodpecker (*Campephilus principalis*), even though there have been no new confirmed sightings of the bird.

Since a group led by ornithologists at Cornell University made the first claimed sighting for more than 50 years last spring, teams have unsuccessfully searched the Big Woods region of Arkansas for confirmation (see *Nature* 437, 188–190; 2005). Over the past winter, birdwatchers reported a half-dozen possible sightings, but there have been no photographs or other solid evidence. Sceptics of the discovery dismiss the sighting claims as "faith-based ornithology".

The budget request to Congress calls for \$1.6 million to develop the recovery plan for the bird; \$400,000 for searching the lower Mississippi River Valley; and \$200,000 for law-enforcement support.

Dutch universities ditch reported Nazi collaborator

The University of Utrecht in the Netherlands has dropped 1936 chemistry Nobel laureate Peter Debye from the name of its nanomaterials institute. The move follows reports that the former national hero had actively supported Nazi policy in the 1930s.

In addition, the University of Maastricht has dropped Debye's name from its international science prize, which is funded by the Edmond Hustinx Foundation. The university is also now planning to produce a monograph investigating the life of Debye.

The reports of Debye's Nazi links were made in *Einstein in the Netherlands*, a Dutch-language book published in January. Science writer Sybe Rispens found evidence suggesting that Debye had, for example, expelled Jews from the Kaiser Wilhelm Institute for Physics in Berlin when he was its director.

Institute that cloned Dolly heads for pastures new

The centre that created Dolly the sheep could be set to close.

The Roslin Institute near Edinburgh, UK, made the headlines when its staff unveiled Dolly (pictured) in 1997. The creation of the world's first mammal to be cloned from an adult cell pushed debate about the ethics of cloning onto the public stage.

Under plans announced on 24 February, much of the institute's science will merge with efforts at other local research organizations and move to a new facility, the Edinburgh Bioscience Research Centre, largely funded by the Biotechnology and Biological Sciences Research Council. The future of the Roslin site is yet to be decided.

Roslin's director, Harry Griffin, has campaigned for the move, saying the merger will save costs and create a critical mass of researchers. The move could be complete by 2009, he says.

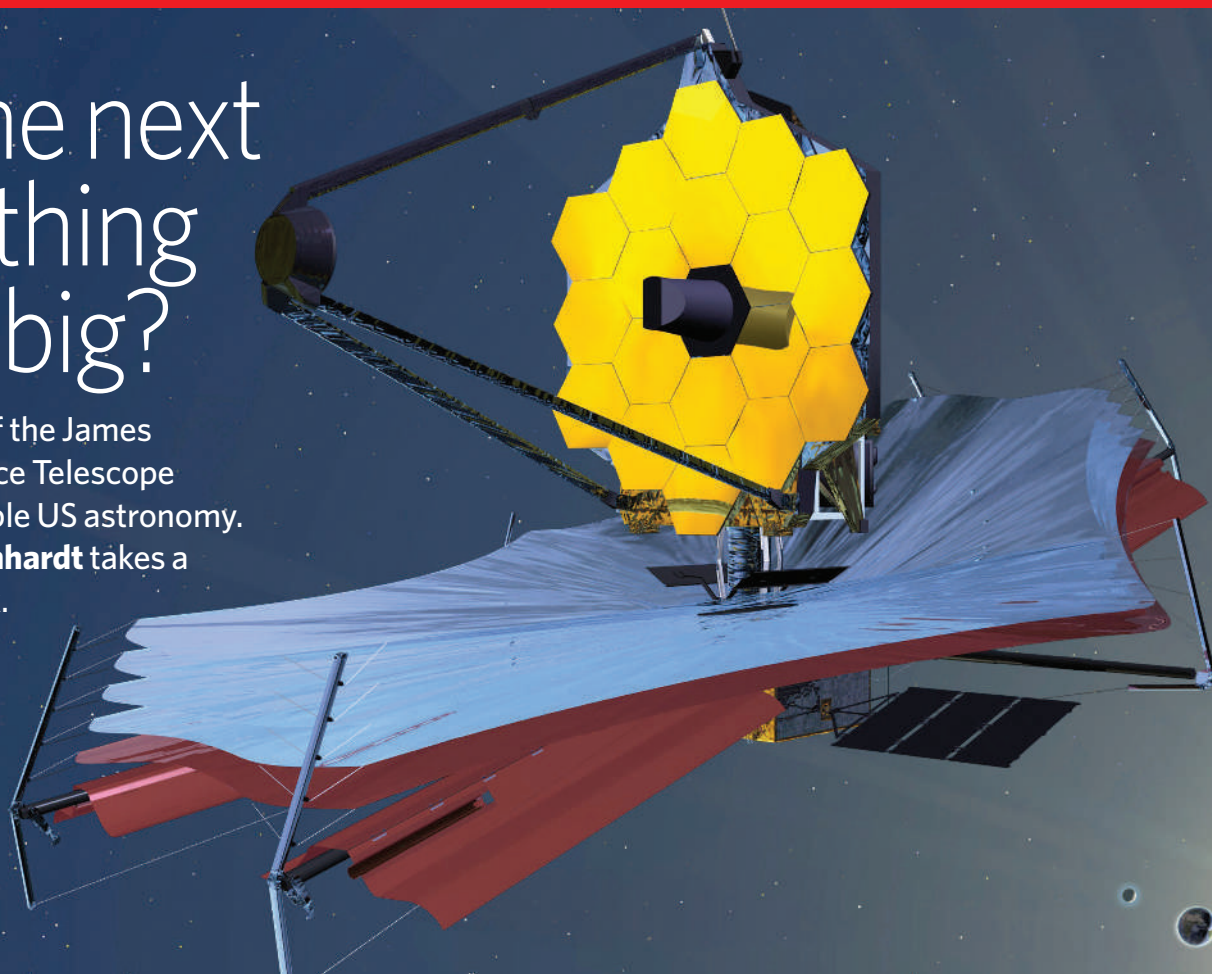


REUTERS/J. J. MITCHELL

M. GODFREY/THE NATURE CONSERVANCY

Is the next big thing too big?

The cost of the James Webb Space Telescope could cripple US astronomy. **Tony Reichhardt** takes a closer look.



Every ten years, US astronomers get together to list the things they want most in all the world — and outside it. This may sound like a grown up version of writing to Santa Claus, but these Decadal Surveys are taken very seriously. Presenting a united front on what matters most to one's profession is a powerful bargaining tool when projects come up for political approval.

On astronomers' most recent wish list, put together in 2000, pride of place was given to what was then known as the Next Generation Space Telescope, an observatory that would take up the mantle of the Hubble Space Telescope as Earth's orbiting eye on the cosmos. Half a decade on, that telescope, now named after former NASA administrator James Webb, is well under way. But, as always, there's a catch. At the beginning of the millennium, US astronomers thought that their most-wanted project would cost \$1 billion. Its projected cost is now nearly five times that.

Price tags that mimic the Big Bang's inflation are nothing new to astronomy. The problem for the James Webb Space Telescope (JWST) is that the budgetary space in which it's expanding is shrinking. Money once slated for science is being diverted to the space shuttle, the International Space Station and plans for future

manned exploration (see *Nature* 439, 768–769; 2006). So the costs of the Webb telescope are leading to cancellations — or 'indefinite deferrals', as NASA prefers to call them. For those whose dreams are crushed in this process, the Webb telescope is looking less like the future of their field and more like its foreclosure. One critic of the process is Shri Kulkarni, an astronomer at the California Institute of Technology in Pasadena. "My worry," he says, "is that we are starting on a project whose cost we don't understand, and which is now devouring space-based astronomy. I doubt there's any project that is worth abandoning the rest of the field."

"We are starting on a project whose cost we don't understand, which is devouring astronomy."

— Shri Kulkarni

Sterl Phinney is an astrophysicist, also at the California Institute of Technology, whose favoured project, the LISA gravitational wave telescope, has just been deferred indefinitely by NASA. He says that even before the recent cuts, the Webb telescope was "basically sucking up all the other money" astronomers hoped to use. The community is now split between those who view the situation with growing alarm and those who, according to Phinney, "really like the JWST and think it's OK that it eats everything else — although even some of those are worried about the balance and health of astronomy." Kulkarni thinks some in this second camp are still in "stunned shock"

over the most recent shift of funds away from science at NASA, and just haven't reached a consensus on what to say, let alone do, about it.

One thing on which everyone agrees is that, if it works as advertised, the Webb telescope will be one fantastic machine. The telescope's 25-square-metre mirror is not just much bigger than Hubble's; it is bigger than any you would have found at any observatory in the world when Hubble launched.

The long view

Hubble's design is optimized for visible and ultraviolet light, but the Webb telescope will see in the infrared. Sitting above the atmosphere, it will have an unfiltered view of a swathe of wavelengths from 0.6 μm (at the red end of the visible spectrum) all the way to the first fringes of the far infrared at 28 μm . At longer wavelengths, images of a given resolution require a larger mirror; the Webb telescope's honeycomb of burnished beryllium will give it a resolution in the infrared that is as sharp as Hubble's is in the visible. The mirror's size also makes the telescope particularly sensitive: its instruments should see objects 10 to 100 times fainter than Hubble can.

Going into the infrared means the telescope has to have a big mirror and has to be stationed far from Earth (the heat from which would otherwise be a problem). It also has to be thoroughly shielded from the Sun, with a structure that somewhat resembles a multi-

NORTHROP GRUMMAN

storey trampoline. These requirements have all driven up the telescope's cost. But seeing in the infrared is not an optional extra; it's a necessity. If you want to look at the early Universe, the infrared is where the action is.

Theory holds that after the glow of the Big Bang faded, the Universe entered a long, lightless 'dark age'. Eventually, knots in the cold dark material condensed, collapsed and began to shine — the first stars. These earliest stars are receding from us at a great rate, which stretches out the light that reaches us and extends its wavelength towards the red end of the spectrum. The first stars are thought to have started shining less than a billion years after the Big Bang, giving them 'redshifts' in which the change in the light's wavelength relative to its original value is 20 or more — moving visible wavelengths well into the infrared. This is a large part of the reason why even far-sighted Hubble has never seen objects with a redshift of more than 7. The Webb telescope should solve this: young stars characteristically give off ultraviolet light that, after a redshift of 15, shines at $1.9\ \mu\text{m}$ — "smack in the middle of our best band", enthuses John Mather, the JWST senior project scientist.

First light

Recently, Mather was part of a team led by Alexander Kashlinsky, now of Goddard Space Flight Center near Washington DC, that used the much smaller Spitzer infrared space telescope to detect a diffuse glow from 'first light' stars (A. Kashlinsky *et al.* *Nature* **438**, 45–50; 2005). No current or planned telescope, not even the Webb telescope, can resolve individual first-light stars. But the Webb telescope should be able to see the supernovae that resulted when these massive but short-lived bodies exploded, providing the Universe with its first heavy elements.

It should also see the first galaxies that formed. One of the key observations for the Webb telescope will be 'deep field' pictures similar to those taken by Hubble. In these, a telescope points at a small patch of sky, taking a long, deep exposure that is designed to reveal extremely faint, distant objects. Astronomers hope the Webb telescope's near-infrared deep field (where contrast and resolution are best) will provide them with images of the very first galaxies and proto-galaxies.

For all this, advocates of the Webb telescope are eager to point out that it is more than just a 'first light' machine. They argue — especially to scientists whose projects are being sacrificed — that although the Webb telescope was inspired by cosmologists' interest in the earliest stars, it has much to offer other fields.

Take, for example, the search for planets around other stars. One of the casualties of this year's NASA budget was the Terrestrial Planet Finder, a mission designed to look for objects the size of Earth; its budget fell to zero. The Webb telescope cannot do what the Terrestrial Planet Finder was meant to do. But

planetary scientist Jonathan Lunine of the University of Arizona, Tucson, points out that it should still deliver relevant science that no current telescope can. An interdisciplinary investigator on the telescope's science working group, Lunine says it will return images and spectra for planets not all that much bigger than Jupiter, and may in special circumstances produce spectra for the atmospheres of planets as small as Uranus. Its high-resolution pictures of dusty circumstellar disks will be the sharpest ever, providing insight into planet formation. It even has applications within our own Solar System, for studying the thermal properties of the Kuiper-belt objects that orbit beyond Neptune.

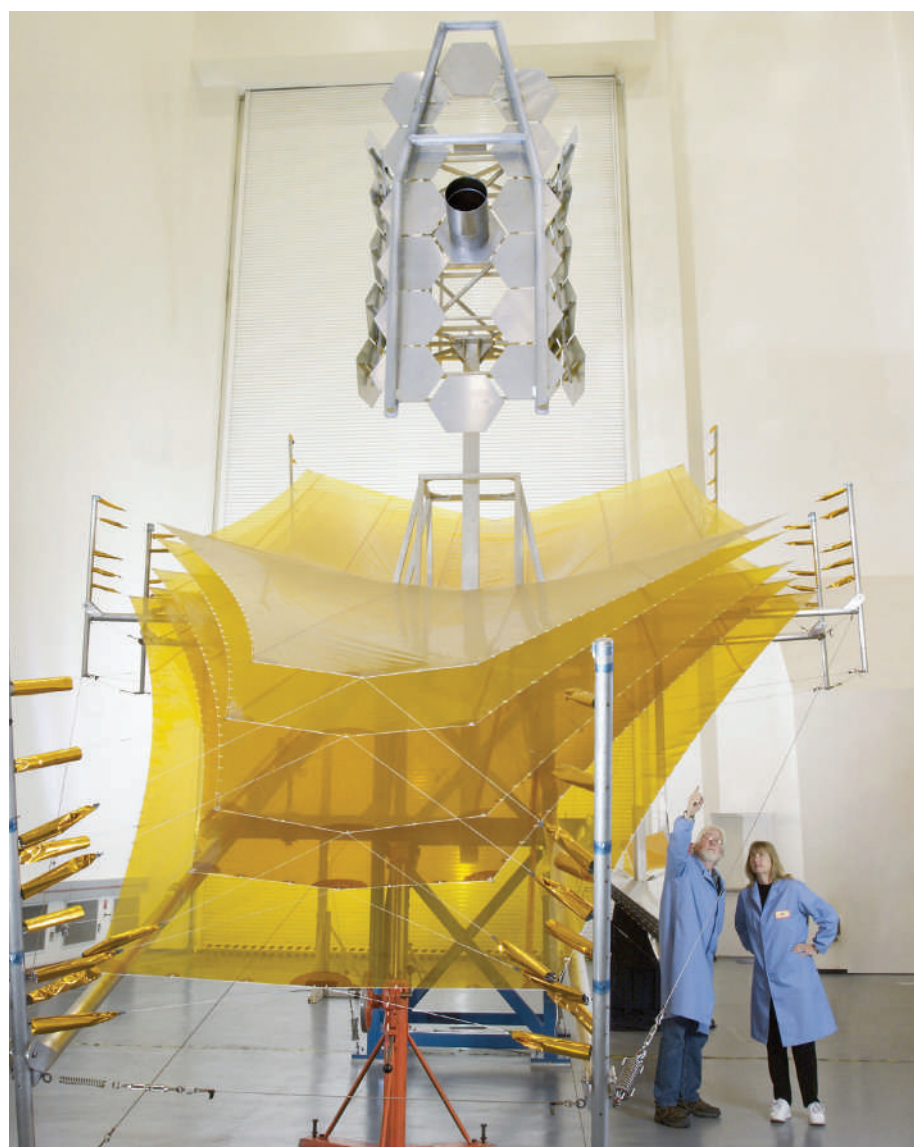
And these are just the planned observations. Heidi Hammel, a planetary astronomer at the Space Science Institute in Boulder, Colorado, and another member of the JWST science working group, says some of the telescope's most important results may well be unforeseen. Some of Hubble's best findings, including the deep-field observations, "came from things

we hadn't even thought of, because it opened up new discovery space", she says.

Cash register

So no one is denying that the JWST will be a first-rate telescope, perhaps even a revolutionary one. Just last August an independent assessment team charged by the project to review the telescope's science potential reported that "the scientific case for the JWST mission has become even stronger" since the Decadal Survey's endorsement in 2000. But what of its expense? NASA's latest budget puts the project's price tag, including \$1 billion for a decade's worth of operations, at \$4.5 billion. That's more than the entire annual research and development budget of the National Science Foundation; it represents more than \$1 million for each full member of the American Astronomical Society.

On top of that there are the contributions by Europe and Canada, junior partners on this telescope just as they were on Hubble. Europe will contribute one of the telescope's four



Sun screen: this half-scale model shows how the Webb telescope will be shielded from solar heat.

NORTHROP GRUMMAN

instruments, the Near Infrared Spectrograph, and the launch on an Ariane 5 rocket. For an investment that approaches half a billion dollars, it will get 15% of the viewing time. Canada's \$57 million will provide a fine guidance sensor and other hardware, for which it gets about 5% of the science use.

The total cost is more than 30 times greater than that of the Keck telescopes on Hawaii, which boast two of the largest mirrors on Earth. One reason for this extraordinary expense is that the JWST is a challenging spacecraft to build. The segmented structure of the mirror, made from 18 hexagonal pieces of beryllium, is unlike anything built before; so is the multilayer sunshade and the system that will deploy them both. Robert O'Dell, who as project scientist for Hubble was in Mather's position 30 years ago, points out that Hubble was able to borrow much of its technology from spy satellites. The Webb telescope has no such heritage on which to draw.

Astronomical costs

NASA tried to head off difficulties by tackling some key technology issues early in the project's life. Although that helped to identify potential trouble spots, it didn't always reduce costs, says Eric Smith, programme scientist for the Webb telescope at NASA headquarters in Washington DC. For example, the engineers found they could build lightweight mirror segments, but not as fast as some had hoped — the job will end up taking six years instead of four. The early development work led to the mirror losing a third of its originally envisaged surface area in 2001. Other proposed cuts in capability — the dropping of the telescope's mid-infrared instrument, a possible further shrinkage to the mirror — were deemed scientifically unacceptable.

Some savings have been found. In 2005, project managers decided to forgo the extra mirror polishing needed to make the telescope's images utterly crisp in wavelengths shorter than

1.7 μm , on the basis that future large ground-based telescopes equipped with adaptive optics would be able to deal with these wavelengths more-or-less as well. And switching from a vacuum test chamber in Ohio to one at the Johnson Space Center in Houston, Texas, should save more than \$100 million.

Despite this, the cost has continued to climb, alarmingly jumping almost \$1 billion in 2005 alone (see graph). NASA's requirement that the programme beef up its contingency fund added a little over \$200 million. A delay in the government's decision to move from a US launcher to the Ariane added an estimated \$300 million as highly paid engineers were unable to move forward until they knew which rocket they were designing for. The situation is particularly embarrassing given that the cost of delaying the decision ended up being greater than the cost of the launch. That delay, and a NASA decision to rearrange the project's long-term budget yet again, saw the launch slip from 2011 to its current date of 2013. Every slip increases the total cost.

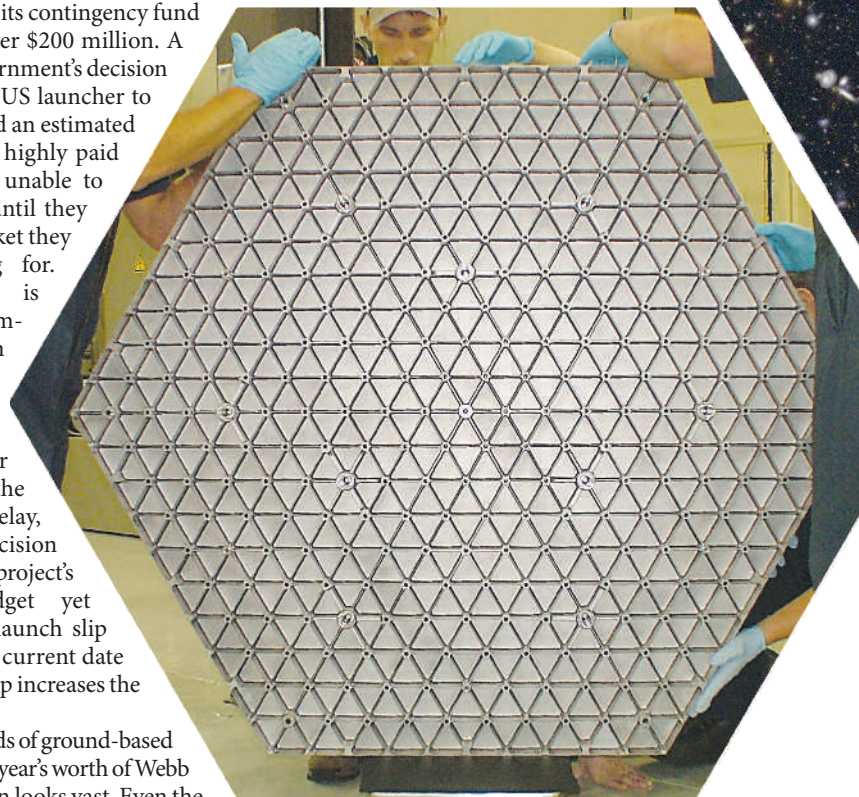
By the standards of ground-based astronomy, just a year's worth of Webb telescope overrun looks vast. Even the most expensive proposed instruments, such as the Atacama Large Millimeter Array (see *Nature* **439**, 526–528; 2006) or the various Keck-dwarfing 30-metre telescopes that are under discussion, should leave ample change from \$1 billion. But O'Dell offers some perspective. Space telescopes are more expensive not just because the technology is more challenging, but because every problem and every contingency has to be thought through and solved before launch. This typically requires a large team of engineers to remain in place for years. What seem to be additional costs have also come from NASA's long and painful switch to 'full-cost accounting'. In this system, all of a mission's expenses — every paper clip and every guard at the front gate — are included in the total

bill. This makes NASA overheads smaller, and the prices of individual missions greater.

For all this, the growth in cost of the Webb telescope is not unprecedented. O'Dell recalls that in 1972, Hubble's total price including its first year of operation was projected to be \$300 million (\$1 billion in today's prices). According to Robert Smith, a historian at Canada's University of Alberta who wrote a political history of the telescope, Hubble ended up costing a lot more by the time it reached the launch pad in 1990 (several years late owing to the Challenger shuttle

"It's like rush hour traffic: suddenly everybody's piled up, there are delays, and it becomes unstable."

— David Black

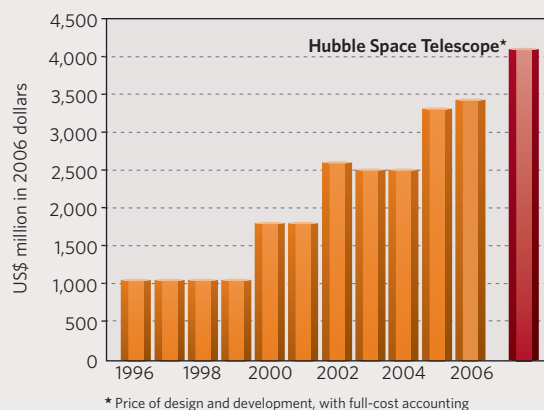


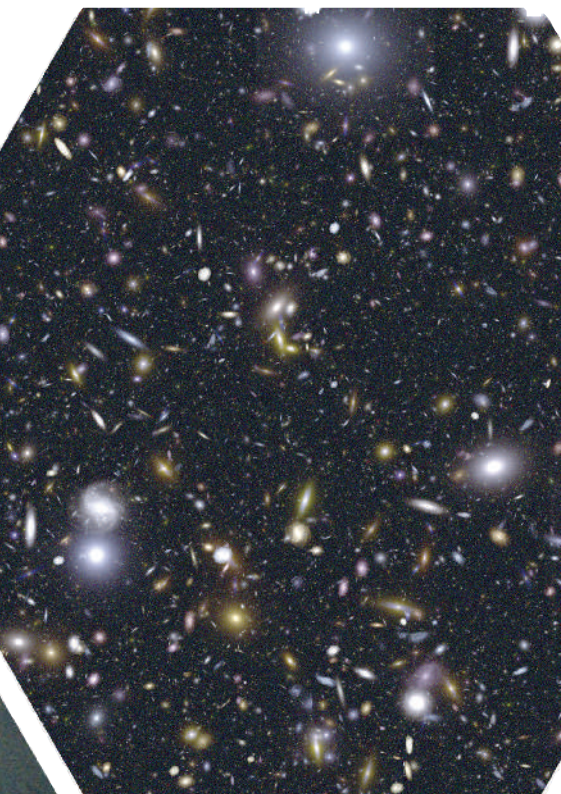
accident). He says that if the budgets were calculated according to NASA's current full-cost accounting standards, "the development cost of Hubble to date is certainly more than \$4 billion in today's dollars".

NASA's Eric Smith adds that when new instruments and operating expenses are added, that comes to \$9 billion. This doesn't include the cost of four space-shuttle servicing missions to Hubble, and a fifth being planned — the cost of a shuttle launch can be put at about \$500 million. All in all, building, launching, using and refurbishing Hubble has probably been the most expensive undertaking ever made in the name of pure science; the mission is still, remarkably, costing more than \$300 million a year.

In that context, you begin to understand how Lunine can claim with a straight face that the Webb telescope — which will outperform Hubble in almost every way — is in fact "a bargain" at \$4.5 billion. Still, the discipline as a whole has to wonder whether it can afford a

PROJECTED COST OF THE JAMES WEBB SPACE TELESCOPE





Far out: the Webb telescope will have an array of hexagonal mirrors (left) that work to find the most distant galaxies yet.

bargain quite this big in straitened times. Between them, Hubble and the Webb telescope will soon consume half of NASA's astrophysics budget (see chart).

Some critics have concluded that the carefully crafted recommendations of the most recent Decadal Survey are no longer viable: if the costs had been clear, the priorities might well have been different. Kulkarni was on the review panel for ultraviolet, optical and infrared astronomy from space. He says, "I now regret that we were not clear thinkers" about what was affordable, and he believes that the plan was "fiscally unrealistic" even before NASA cut its science budget last month.

If it was unrealistic, says David Black, president of the Universities Space Research Association, scientists should share in the blame. "Astronomers pushed NASA to have all these missions. NASA bought into that. And all it

takes is one hiccup like the JWST overrun. It's like rush hour traffic. One incident, and suddenly everybody's piled up, there are schedule delays, and it becomes unstable very quickly."

The problems that come with sending mission after mission into this crowded traffic are exacerbated when the costs of the missions are set artificially low at the beginning. When NASA administrator Mike Griffin told a January meeting of the American Astronomical Society that the Webb telescope wasn't so much overbudget today as it was "undercosted" at its inception, he wasn't just putting a good spin on things.

The Decadal Survey guessed the cost as \$1 billion. Studies in the mid-1990s had pegged the price at between \$500 million and \$1 billion. These were based partly on the hope — unfulfilled, as it happened — that the Webb telescope might take advantage of advances in building low-cost spacecraft developed by the military. Oddly, earlier cost estimates for a large infrared space telescope were closer to the mark. A 1984 Space Science Board panel predicted the cost including operations to be \$4 billion (roughly \$7 billion today), and a subpanel of the 1990 Decadal Survey thought it would run to about \$2 billion not counting operations, which, when adjusted for inflation, closely matches NASA's current projections.

Political plays

Garth Illingworth of the University of California, Santa Cruz, who chaired the 1990 panel, chalks the anomalously low estimates from the 1990s up to a "lack of reality" inherent in the 'faster, better, cheaper' philosophy of Dan Goldin, NASA's administrator at the time. Goldin focused on accelerating the development of spacecraft, and increasing innovation, while accepting a moderate rise in the risk of failure. Some projects conceived under this tag, in particular two Mars missions lost in quick succession, brought it a certain disrepute. "It was a horrible, political circumstance framing all the discussion in that decade," says Illingworth.

Reinhard Genzel of Germany's Max Planck Institute for Extraterrestrial Physics in Garching says it was clear at the time that a \$500-million estimate for the Webb telescope was a "political price". Yet such was the climate of the 1990s that when estimates for the European Space Agency's smaller Herschel infrared observatory came in at \$1 billion, he says, "I was approached by many colleagues saying, 'You

guys are so stupid. Why can't you do this for less?' That's now haunting NASA, of course."

Today, Illingworth inveighs against the "extraordinarily bad, artificial cost estimates" of the Goldin era. But the 2000 Decadal Survey seems to have been happy to accept them. The world of big science is well used to projects being lowballed — a process that gets schemes started on the basis of a low cost estimate, with the implicit hope that by the time the true costs are known inertia and vested interests will make it impossible to pull out. Lowballing is not a practice anyone would defend on principle, but histories like the Hubble's show it can work (see page 127).

Past the disquiet

Craig Wheeler, a University of Texas astrophysicist and president-elect of the American Astronomical Society, takes the lessons of Hubble to heart: "I remember when we were building the Hubble Space Telescope, which has been spectacularly successful, there were an awful lot of eggs put in that basket. And other smaller, faster, university-based projects suffered. I think we got through it." Wheeler accepts the disquiet over the Webb telescope's costs, but he doesn't think astronomers have yet reached "the point we collectively would say 'enough'".

And he warns against revisiting the results of the Decadal Survey on the basis of the current crisis: "You alter those priorities at great risk."

Kulkarni is more

pessimistic. He thinks that NASA's "laserlike focus" on the Webb telescope short-changes missions that would hunt for planets, probe the nature of dark matter, search for gravitational waves, and tackle other topics that might ultimately prove more popular with young scientists and with the public.

His concern is shared by Charles Beichman of the Jet Propulsion Laboratory in Pasadena, a leading light of the cancelled Terrestrial Planet Finder mission. Beichman thinks the Webb telescope will be "a fine machine. It will do fantastic science". In fact, he is on one of the instrument teams. But when he goes to professional meetings, he sees more young astronomers attending sessions on planet-finding than on Hubble or the Webb telescope.

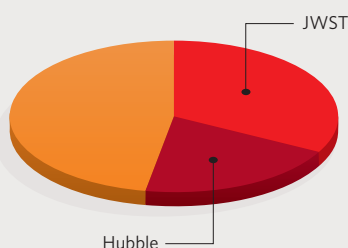
Lunine thinks the critics are fighting the wrong battle (and that anyone who doesn't realize that the Terrestrial Planet Finder would be costlier than the Webb telescope is dreaming). It is not the JWST that is to blame, he says. The real problem is that "NASA's science budget is not adequate", and science is "taking the hit" as the agency shifts its focus to returning astronauts to the Moon. That may be the case. But for now, the Webb telescope is left in the awkward position of being the only one eating in a room full of hungry people. ■

Tony Reichardt writes for *Nature* from Washington DC.

"I remember when we were building Hubble, other projects suffered. I think we got through it."

— Craig Wheeler

NASA'S ASTROPHYSICS BUDGET
US\$1.5 billion in 2008



BALL AEROSPACE & TECHNOLOGIES; NASA

SOURCE: NASA



CAUGHT BETWEEN SHORES

Ecologists paid by industry to assess the effects of businesses on the environment are often accused of selling their souls. But isn't scientific expertise exactly what is needed? **Michael Hopkin** investigates.

Picture this: you are a talented research ecologist and you're evaluating whether a planned hydroelectric dam could damage the local ecosystem. Your findings lead you to believe that the fish in the river would not be significantly harmed by the dam. But when you publish your results, your colleagues refuse to believe them. Why? Because you work for the company that is building the dam.

At first glance, big business seems to be only bad for the environment. After all, industry has cut down rainforests and opened up huge mining scars on the landscape. Surely, it might seem, any ecologist who takes money from an organization that so harms the natural world must be putting concerns about the environment second to salary.

In fact, many ecologists take up industrial contracts to try to minimize the damage caused. But in doing so, they walk a delicate line between those who want to save the natural world and those who want to exploit its resources. Some face accusations from their peers that they've 'sold out'. And conflicts often arise between their interests as researchers and those of the companies they work for. Faced with these challenges, many must question whether their decision to work for industry was worth it.

For some, such as Tyrone Hayes, the answer is no. Hayes is an ecological toxicologist at the University of California, Berkeley, who has received funding from the agribiotech giant Syngenta in the past. "My view has changed a

lot since working with Syngenta," he says. "It's made me a lot more sceptical of scientists who get involved with industry."

Hayes's work touches on one of the most politically charged areas of applied research: the impact of pesticides on the environment. Specifically, he is studying the effects on frogs of atrazine, widely used on transgenic crops. At Berkeley, Hayes took up a contract with

"People I barely knew — students, government regulators, and friends of friends — started questioning my integrity." — Bill Streever

Syngenta, brokered by a consulting firm. In his research, he found that exposure to atrazine leads to male frogs becoming feminized, as measured, among other traits, by larynx size¹.

Culture of mistrust

But Syngenta asked him to divide his data on larynx size by the frogs' body weight, a procedure that he says was designed to make the key finding disappear. A Syngenta representative said that processing the information in this way is a common method for handling data from such studies, designed simply to control for the presence of naturally stunted frogs. Hayes eventually gave up the lucrative contract, and no longer works with industry colleagues, who are forbidden from discussing their results with him.

Hayes worries that many scientists in his field

could come under similar pressure when working under industry contracts. "It's up to researchers to maintain the integrity to say: 'No, I produce data in my lab and I have got to stand by them,'" he says. But that can be hard to do, especially when research funding is at stake.

In general, big businesses, often forced by environmental regulations to investigate the impact of their activities on the environment, have the financial muscle to fully support such projects. Last year the BP Conservation Project, funded by the UK-based oil giant BP, awarded US\$600,000 to 28 groups of conservationists. Overall, the company spends around \$100 million each year on community-investment efforts. That amount of money can seem vast to researchers used to relying on academic funding (see 'The lure of industry').

But peers remaining in academia often charge that industry employees have sold their integrity down the river. "When I went to work for 'big oil', friends wondered why I had gone over to the dark side," recalls Bill Streever, who works for BP as an environmental investigator in Alaska. In a journal editorial published last year², Streever wrote: "People I barely knew — students in lecture halls, government regulators, and friends of friends — started questioning my integrity."

Asked whether he accepts that there is conflict between the work he does and the source of its funding, Streever's answer is an emphatic no. Working inside industry, he argues, is the best way to gain some real

G. LUDWIG/VISUM/PANOS

influence and ensure that big business really does embrace environmental stewardship.

Many companies need competent scientists to meet legal guidelines on minimizing the detrimental impact of their operations. "My feeling is that there's a tremendous need out there for talented scientists," says Streever.

Streever also enjoys the relative largesse of big business. For example, in an investigation he mounted on whether noise from oil-drilling operations affects whales in the Beaufort Sea, Streever used BP money to convene a panel that brought together expertise from the diverse fields of acoustics, statistics and mammal ecology. As a wetlands ecologist by training, Streever says it would have been difficult to obtain the funding needed for the panel through traditional wetlands-ecology channels.

Mutual gains

Scientists aren't the only ones who stand to benefit from partnerships with industry. Elaine Dorward-King, head of health, safety and the environment at the mining multinational Rio Tinto, argues that companies have everything to gain by protecting their reputations through diligent environmental stewardship. "If you want to have a positive impact on the world, I can scarcely think of a better area to work in," she says.

For a corporate behemoth such as Rio Tinto, she adds, losing the trust of local people means losing out financially in the long run. In Madagascar, where the company is eyeing a huge deposit of the mineral ilmenite, Rio Tinto's environmental department has spent the past 20 years weighing up the likely impact of a mining operation on the local community and



P. S. TURNER

Put off: Tyrone Hayes's experience with an agribiotech giant has made him mistrust scientists in industry.

environment. Past mistakes stress this need for caution: in Papua New Guinea, back in the 1960s, a copper mine owned in part by Rio Tinto polluted the surrounding environment, and eventually helped lead to a civil war that caused the mine to close.

There is often a third player in the collaboration between environmental science and business: non-governmental organizations,

or NGOs. Many operate from a position of supreme cynicism, accusing industry figures such as Dorward-King of spouting platitudes while having one eye firmly fixed on the balance sheet. But others see collaborations between science and industry as a key opportunity to influence big business, and lever their relatively modest funds into brokering such collaborations.

One organization that falls into the second category is Fauna and Flora International, based in Cambridge, UK. It joined forces with BP in 1990 to help minimize the impact of oil-drilling operations on local wildlife, and was predictably accused by other, more hardline NGOs of selling out.

But Annelisa Grigg, the organization's director of corporate affairs, argues that this puts it in a position of influence with big business that it would not otherwise enjoy. She says that her organization now has the ear of more than 25 mining corporations, as well as the International Council of Mining and Metals, based in London, UK.

In the end, environmental scientists say they must keep a critical eye on all science, whether it comes from the industrial or academic sector. "You always have to prove yourself," says Marianne Carter, who manages a conservation programme that links BP with conservationists in academia. "And show that it's not just a greenwash."

Michael Hopkin is a reporter for news@nature.com.

The lure of industry

Do environmental scientists make more money in industry than in academia? Most people would probably say yes, but the picture isn't so clear-cut. Published salary data suggest that the average contract scientist in the United States takes home a similar pay cheque to a university instructor. And for those who scale the ranks of academia, the pay offs for university employees can be even higher.

The Commission on Professionals in Science and Technology (CPST), which produces average-salary data for US scientists, shows that in 2004-05 the average rank-and-file university academic working on disciplines related to natural resources and conservation

earned almost US\$41,000. The average figure for academic ecologists fell short of this benchmark, at around \$32,600. Meanwhile, environmental scientists working as consulting staff researchers to the corporate sector earned slightly more than \$39,000.

Elaine Dorward-King, who oversees environmental surveying for the mining multinational Rio Tinto, doesn't believe that money is a big motivator, because those in search of big bucks would be pursuing different careers anyway. "If you want to get rich, you don't become an environmental scientist," she says. In fact, she argues, the way to rake in the cash is to stay on the academic tenure track. "Full professors

make substantial salaries," she says.

Does she have a point? Possibly. The CPST data show that US professors of ecology bring home on average more than \$90,000. But this is surely small beans compared with what someone with similar management responsibility would earn in the corporate sector.

And what's more, pay cheques in US academia are widely regarded as healthier than those at universities in other countries. So although US scientists may be tempted into industrial positions by factors other than money, the financial enticement may be stronger for those in other countries. **M.H**

1. Hayes, T. B. *et al. Proc. Natl Acad. Sci. USA* **99**, 5476-5480 (2002).
2. Streever, B. *Frontiers Ecol. Env.* **3**, 407 (2005).

BUSINESS

Ringling the changes at Bell Labs

Lucent hopes that a leader with an entrepreneurial bent will revive the legendary Bell Laboratories. Geoff Brumfiel reports.

Jeong Kim is having a frustrating day. The president of Bell Labs is supposed to be teleworking from his spacious home in Potomac, Maryland, but his broadband Internet connection is down and the repairman is nowhere to be found. When asked why he doesn't get his own researchers to work on the case, he shrugs: "I hadn't thought of that."

But 45-year-old Kim has been thinking a lot lately about the scientists and engineers that he oversees. Since taking the helm in April last year, Kim has been working overtime to change the culture of the world-famous Bell Labs at Murray Hill, New Jersey. He has reorganized researchers into smaller, interdisciplinary teams whose projects compete with each other for lab support. He has increased dialogue between researchers and business managers by arranging in-house technology summits. And he has dramatically reduced the number of technology-development projects, making sure those that remain are closely tied to company needs.

Fresh face

For the legendary 81-year-old lab, this Korean-born reliability engineer represents a break with the past. Most of Bell's previous presidents spent years researching there; Kim is an outsider. He joined Bell's owner, Lucent Technologies, in 1998 after the firm bought up his Maryland-based voice and video networking company, Yurie Systems, for \$1 billion. He has chosen not to move to Murray Hill, but instead commutes the 300 km from his Maryland home for four days a week.

Lab scientists and outside observers say that Kim's appearance has created fresh enthusiasm at Bell, which was once one of the world's greatest research centres but has fallen on hard times in the past decade. "Jeong brings this energy," says Nobel laureate Horst Störmer, a physicist at New York's Columbia University who was at Bell in its heyday and still works there part-time. "He's an entrepreneur; he's a different kind of person from previous leaders."

Patricia Russo, Lucent's chief executive, says that Kim is teaching Bell's academically minded researchers to work better with

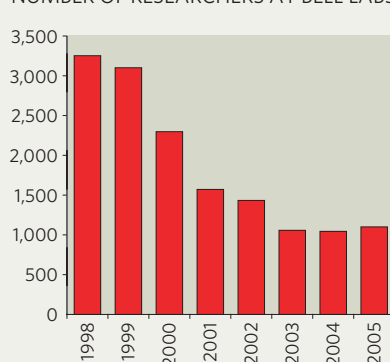
business executives. "I think it's fair to say that under Jeong Kim's leadership there is tremendous alignment," she says.

But changing course will not be easy. Through much of its history, Bell's cosy lab benches have resembled a university physics department more than an industrial research facility. Bell Laboratories was originally formed as the research arm of AT&T, a company that held a monopoly on US telecommunications for half a century. AT&T saw the lab as a brains trust: it employed thousands of scientists there, and allowed them to pursue a wide variety of interests. The lab became as well known for its Nobel-winning research — such as the 1965 discovery of the Big Bang's afterglow — as it was for commercial innovations such as the C programming language.

The halcyon days ended with the telecommunication industry's deregulation in the late 1980s. But the real shock came when AT&T spun off Lucent in 1996, according to Robert Calderbank, an electrical engineer at Princeton University and former AT&T vice-president for research. He says that the new, smaller company wasn't sure at first what to do with the lab.

For a while, Calderbank says, it trumpeted Bell's achievements — but relied on the acquisition of other companies for new technology. "At some point, I think the brand 'Bell Labs' became more important to Lucent than the technology." Things went from bad to worse when demand for telecommunications equip-

NUMBER OF RESEARCHERS AT BELL LABS



Bell's new chief, Jeong Kim, is looking ahead.

ment collapsed in 2001, leading Lucent to spin-off certain lab functions and downsize others.

Kim says he has made it his mission to improve morale and strengthen the relationship between Bell and its parent company. "I've clearly stated that we're going to be successful by making Lucent successful," Kim says.

Focus on the useful

"We're trying to create more of an entrepreneurial spirit internally," agrees David Bishop, Bell's vice-president for nanotechnology research. He points to Kim's reorganization of the lab into teams that resemble small start-up companies and compete for funding. Bishop says that his nanotechnology group is now tightly focused on developing devices that improve communication over long distances — possibly by enabling remote touch or smell.

Basic research that does not feed into Lucent products immediately will still go on, Kim adds. But now researchers are being encouraged to find external sources of funding for such work. The lab's quantum-computing group, for example, is working with money from the Pentagon's Defense Advanced Research Projects Agency.

In the arena of development, Kim has gone still further. When he came to the lab, he says, the business unit and lab researchers were working on dozens of costly projects, with few clear goals. So he created a formal prioritization process. As a result, the number of projects



M. TEMCHINE

has been cut from 160 to 30, saving money and freeing up personnel for other tasks.

These changes are viewed with scepticism by some Bell veterans. "I think start-up firms are much better than large telecom firms at taking chances," says Peter Shor, a mathematician formerly at AT&T, and now at the Massachusetts Institute of Technology. He adds that Kim's plans for creating a competitive environment could discourage researchers from pursuing some interesting lines of inquiry.

Others take the view that Lucent is just too small to use a lab the size of Bell. In recent years, most of its business has come from maintaining the equipment it has already installed, points out John Slack, an equity analyst with Morningstar, a Chicago financial-research firm. "Bell Labs doesn't have a huge impact on Lucent's operating business," he says.

Slack adds that however competitive the lab becomes, its fate ultimately depends on its parent. After several divestments, Lucent's revenues fell from \$34 billion in 2000 to just under \$10 billion last year. Revenues are now growing, but until Lucent gets on a firmer financial footing, Bell is unlikely to see a major rebound, Slack says.

Kim accepts these points, but remains bullish about the lab. This year, it increased its total research staff for the first time in nearly a decade, he points out (see graph). "There are smart people everywhere, but what's so different about us is our breadth and depth," he says. "It's not whether we do great research and development: it's how we do it."

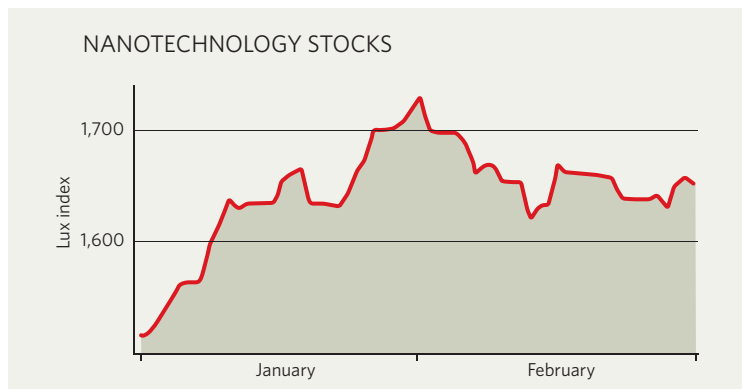
IN BRIEF

VENTER VENTURE Biologist Craig Venter has enlisted a high-profile ally in his quest to build microbes that are specially tailored to solve environmental problems. Ari Patrinos, who has directed biological and environmental research at the US Department of Energy since 1995, was last month appointed president of Synthetic Genomics, a company that Venter founded last year to create such organisms. "I've known Craig for many years and we have always worked very well together," says Patrinos, adding that Venter approached him about the job — based in Rockville, Maryland — several months ago. "I'm going to be 59 this year," says Patrinos. "I felt I wanted to do one other, different thing."

STRONG PRESCRIPTION A US physicians' group has called for special government incentives to push companies into developing antibiotics for drug-resistant hospital nasties such as *Staphylococcus aureus*, better known as staph. In an article in the March issue of *Clinical Infectious Diseases*, the Infectious Diseases Society of America argues that few antibiotics are being developed because they are used for only a short time and do not fetch the prices that give manufacturers a decent return. The society calls for tax breaks and other incentives to get things going.

BOTANIC YIELD A UK biotechnology firm whose drug candidates are based on Chinese plants has launched itself successfully on London's Alternative Investment Market (AIM). Oxford-based Phynova raised about £4 million (US\$7 million) in its share offering on 27 February: two days later its shares were selling for £1.20 — 50% above the offer price. Phynova has six drug candidates under development, including potential therapies for hepatitis C and cancer.

MARKET WATCH



Stocks in companies with a stake in nanotechnology started the year with a bang — although prices fell back a little in February, along with the rest of the market.

The Lux Research nanotechnology index tracks a mixture of companies that sell nanotechnology equipment and products based on nanotechnology, as well as some larger corporations that expect to make use of the technology.

Peter Hebert, president of Lux Research, says that companies performing well so far this year include Nanophase Technologies, an Illinois-based company that makes nanocrystalline materials. Nanophase is one of the oldest companies in the sector, having originally listed on the Nasdaq in 1997, and it has had a turbulent history. But this year its stock rose from \$5 to about \$7 after news

that it is finally managing to sell to customers in the chemicals industry, including BASF.

Other strong performers so far in 2006 include Altairnano, a Nevada materials firm whose stock price has almost doubled since the beginning of January, and Arrowhead Research of Pasadena, California, which invests in start-up companies.

It has been a less happy couple of months for Accelrys, a San Diego-based computer-modelling company whose shares lost about a quarter of their value after the Securities and Exchange Commission asked it to revisit its 2005 accounts.

But Hebert says that, overall, interest in the nascent industry sector is strong: "Industrial corporations are getting more and more active" in nanotech, he says.

Colin Macilwain

Hasty energy review risks failing to win public trust

SIR — The UK government recently announced the launch of its second energy review in three years. Many have commented on the substantive issues and technologies; we wish to draw attention to the limitations of the review process itself. Our concerns draw on three decades of study of the governance of science and technology, and we recognize that, although the review encompasses complex issues involving high stakes and disputed values, it supports a limited degree of public engagement.

Three months have been allocated for decisions that may have implications for thousands of years. Although there is urgency to this issue, and legislative timetables create potential time constraints, the length of the decision-making process should be determined by the scope and complexity of the problem, and the public should be given sufficient notice and resources to respond meaningfully.

An authoritative and legitimate process must be open to a range of possible outcomes, but there is a widespread perception that the government has made its mind up in favour of nuclear energy. If the consultation is to have integrity and to generate trust, the task of initiating the review should be separated from the tasks of convening and implementing it. Furthermore, good review processes must have the authority that comes from participants knowing in advance that their inputs will be taken seriously, and the initiators should be explicit about how the review's findings will be used in future decision-making processes.

The explicit context of such an enquiry, or 'framing' as social scientists call it, is itself contentious. Some may question whether framing the energy-review process in terms of climate-change obligations or the security of gas supplies is appropriate. If a process is to be recognized as legitimate, it must engage participants in the task of choosing the appropriate framing. Scenario approaches like those used by the Intergovernmental Panel on Climate Change to think about alternative policies can produce a manageable number of distinct, internally consistent and plausible alternatives.

Critics of earlier consultation efforts have focused on whether the methods used were representative. We argue that there is no single correct method: there is strength in using an array of tools to address different aspects of the problem and to allow diverse forms of input. We would also urge the government to recognize that non-specialist forms of knowledge, based on people's everyday experiences, can provide a valuable complement to officially recognized expertise. Without substantive public engagement, the

government risks reinforcing the mistrust, controversy and technological failure that we have seen all too often in recent years.

James Tansey

James Martin Institute for Science and Civilization,
Saïd Business School, Park End Street,
Oxford OX1 1HP, UK

Other signatories of this letter:

David Gee *European Environment Agency*

Melissa Leach *Institute of Development Studies, University of Sussex*

Jerome Ravetz *James Martin Institute for Science and Civilization*

Alister Scott *The Knowledge Bridge, East Sussex*

Andrew Stirling *Science and Technology Policy Research,
University of Sussex*

Bronislaw Szerszynski *Centre for the Study of Environmental Change,
Lancaster University*

Tom Wakeford *Policy, Ethics and Life Sciences Research Centre,
University of Newcastle upon Tyne*

Silence isn't necessarily an admission of guilt

SIR — The reaction of C. R. McMahon and colleagues (*Nature* **439**, 392; 2006) to your News story "Animal-rights group sues over 'disturbing' work on sea lions" (*Nature* **436**, 315; 2005) — pointing out that much animal handling, such as seal branding, is a necessary prerequisite to effective conservation — is exceptional, because it represents a rare instance of a voiced scientific opinion on a sensitive topic.

Scientists themselves are often 'branded' by the media, but are ill-equipped to defend themselves, having neither the time nor the schooling in this area. Alternatively, they may fear intimidation or be obliged not to enter the fray by governing bodies.

Of the various groups involved in animal-rights issues, those taking an accusatory stance would do well to consider that silence is not always an admission of guilt.

Rory P. Wilson

Institute of Environmental Sustainability,
School of the Environment and Society,
University of Wales Swansea, Singleton Park,
Swansea SA2 8PP, UK

GOOS can help to keep an eagle eye on the oceans

SIR — Your Editorial "Circulation challenge" (*Nature* **439**, 244; 2006) calls for ocean observations to be "sustained for much longer periods than foreseen in the six-year RAPID programme" and "augmented globally". But you do not mention the Global Ocean Observing System (GOOS; see ioc.unesco.org/goos), which already exists specifically to take up this challenge.

GOOS provides an operational structure for sustained global observations comprising the oceanographic component of the Global

Earth Observing System of Systems. The open-ocean module of GOOS is designed to describe and forecast the state of the physical ocean in order to help understand and predict weather and climate. Its implementation is achieved by pooling the commitments of individual nations, made through their government agencies, navies and oceanographic research institutions, together with sponsorship from international agencies.

Not long ago, *Nature* called for enhanced support for GOOS ("Making sense of the world" *Nature* **433**, 785; 2005). Heeding this call requires first and foremost that nations support the development of a robust commitments mechanism that can efficiently take stock of national efforts and at the same time increase the resources available for coordination. GOOS will also need to increase its efforts to reach, and cater for, a wider range of contributors and users of the system. GOOS welcomes the circulation challenge.

Keith Alverson

GOOS Project Office, IOC/UNESCO,
1 rue Miollis, 75732 Paris, Cedex 15, France

Giants of physics found white-dwarf mass limits

SIR — In his Essay "The death of a star" (*Nature* **438**, 1086; 2005) Freeman Dyson pays a well-deserved tribute to astrophysicist Subrahmanyan Chandrasekhar's lifetime of work. Your readers may be interested to know that Wilhelm Anderson of Tartu University in Estonia and Edmund Stoner of the University of Leeds in England separately published white-dwarf mass limits that predated what has come to be known as the 'Chandrasekhar limit' (S. Chandrasekhar *Astrophys. J.* **74**, 81–82; 1931).

Anderson (*Z. Phys.* **1**, 851–856; 1929) made the fundamental conceptual coupling of relativity and quantum statistical mechanics that leads to a mass limit. And Stoner (*Philos. Mag.* **9**, 944–963; 1930) produced a robust calculation from first principles and calculated the mass limit for a uniform-density star. The Stoner and Chandrasekhar mass limits differ by 20%, owing to their use of different stellar-density models. But this difference is small, given that the presence of any mass limit at all was the key breakthrough.

Eric Blackman

Department of Physics and Astronomy,
University of Rochester, Rochester,
New York 14627, USA

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

BOOKS & ARTS

The life of a sage

J. D. Bernal was a multifaceted crystallographer who laid the foundations of molecular biology.

J. D. Bernal: The Sage of Science

by Andrew Brown

Oxford University Press: 2005. 576 pp. £25

Kenneth C. Holmes

J. D. Bernal, who died in 1971, was one of the most influential scientists of the past century. Andrew Brown's fascinating book on this extraordinary man deepens one's appreciation of this powerful and complex personality who, in addition to his science, was involved in many contemporary political and sociological developments.

Bernal was born in Ireland in 1901 to a truly remarkable woman, Bessie. A tall and beautiful Californian, she was one of the first students at Stanford University. A woman of wide interests, she spoke fluent French (acquired with the appropriate accent in New Orleans) and undertook the Grand Tour, on which she met and married Samuel George Bernal, a moderately prosperous Irish dairy farmer. The daughter of a protestant clergyman, she converted to Catholicism when she married. Her son, John Desmond, grew up a devout Catholic, speaking French and English in rural Ireland, with strong family connections to aunts, uncles and cousins in the United States.

His early passion for science led to his being sent away to Bedford School in England, where he was moderately uncomfortable but studied science and read profusely. He was awarded a scholarship to Emmanuel College, Cambridge, where he read mathematics before switching to natural sciences. Cambridge was in ferment. Young soldiers were returning from the First World War, and a strong sense of disillusion prevailed. Bernal spent countless hours discussing politics and much more besides, thereby earning himself the nickname 'Sage'. In his final undergraduate year, as well as deriving the 230 crystallographic space groups using quaternion notation (and taking his degree), Bernal rowed in the Emmanuel first boat, and spent many happy hours punting to Grantchester with Eileen Sprague, whom he married. As Dorothy Hodgkin remarked in her scientific biography, Sage lived life to the full.

In 1923, on the basis of his extraordinary student tour de force with the space groups and a recommendation from Arthur Hutchinson, Bernal was taken on as a doctoral student by William Bragg, director of the Royal



Through the lens of history: J. D. Bernal, photographed in 1949.

Institution in London, and joined his group of outstanding young crystallographers. But Bernal soon realized that measuring the intensity of X-ray diffraction data with a gold-leaf electroscope was not for him. So he developed the X-ray film method, involving a rotation camera. He also invented the Ewald sphere before Paul Peter Ewald published it (it was essential for making sense of the X-ray photos), but with characteristic generosity he insisted on giving all the credit to Ewald. And he developed the methodology that enabled the X-ray crystal-structure determination of complex molecules.

The Bernals were enjoying their short time in London, living in Bloomsbury, and both joined the Communist Party and the Holborn Labour party. But they returned to Cambridge in 1927, when Bernal was appointed lecturer in structural crystallography in Hutchinson's department at the university. Bernal's laboratory, housed in the old Cavendish lab, quickly became an international centre. His work on the structures of steroids was important and, in retrospect, he narrowly missed receiving a Nobel prize.

Bernal started work on the structure of water after meeting R. H. Fowler on a visit to Moscow. He took on Dorothy Crowfoot (later Hodgkin) as a student and then, a little later, Max Perutz; both were to be awarded a Nobel prize. With Dorothy he showed that crystals of pepsin diffract to high resolution if maintained wet, so the molecules must be identical in structure. This simple observation made the prevailing colloid theories of proteins untenable. And Perutz began a 25-year study on the structure of haemoglobin — a demanding theme for a doctoral thesis. With Isidore Fankuchen, Bernal started work on viruses. He maintained that to understand life you needed to know about protein structures, which could be determined by X-ray crystallography. This was essentially the dawn of molecular biology.

In the 1930s, Bernal became committed to marxism. How a man with such a marvellous analytical mind could come to terms with dialectical materialism is still a subject of discussion — it seems to have been an act of faith, a substitute for Catholicism. Apparently, Bernal's epiphany took place at a meeting on the history of science in London in 1931. The Russian delegation, led by Nikolai Bukharin, arrived late and unannounced. They were given an extra morning to air their views and expounded a theory of the history of science that Bernal made his credo.

In 1938 Bernal was appointed professor of physics at Birkbeck College, London, but at the onset of the Second World War he was pressed into service. The politician John Anderson wanted Bernal as a scientific adviser in the Ministry of Home Security "even if he is as red as the flames of Hell". Together with his friend Solly Zuckerman, Bernal started work on operational research, and they gave bombing a quantitative basis. They had to convince the physicist Frederick Lindemann, a government adviser who proposed the carpet-bombing of German cities, that Bomber Command was

W. SUSCHITZKY/NPG P304

making wildly exaggerated claims about the effectiveness of Allied bombing. Later, Bernal, Patrick Blackett and Zuckerman advised against the bombing of German cities because it was a total waste of manpower and resources.

Bernal and Zuckerman were seconded to Louis Mountbatten's team, which was planning D-day. Bernal's great contribution was to chart the Normandy beaches in detail, identifying every rock, mine and weak spot. His charts turned out to be reliable and the tanks and trucks got ashore.

A strong friendship sprang up between Bernal and Mountbatten, who reported: "Bernal is one of the most engaging personalities I have ever known. Perhaps his most pleasant quality is his generosity. This may be why his great contribution to the war effort has not been properly appreciated." Bernal said of Mountbatten: "He had the habit that great commanders have of acting first and thinking afterwards." Late in the war, Mountbatten became supreme Allied commander in southeast Asia, and Bernal joined him in Ceylon on bomb trials for jungle clearance. He found himself working alongside John Kendrew and they fell to talking about the structure of proteins — a conversation that had considerable repercussions. Shortly afterwards, Kendrew turned up in Cambridge to work with Perutz on haemoglobin and myoglobin, and Bernal's ideas were later very influential in the founding of the European Molecular Biology Laboratory.

After the war, Bernal resumed his professorial duties at Birkbeck, setting up the Biomolecular Research Laboratory in 1948. As well as groups working on organic crystals and proteins, he had others working on computers, the structure of cements (buildings and building materials were a life-long interest), and the structure of water. Rosalind Franklin later joined him to start work on virus structure, which she continued with Aaron Klug.

Bernal continued his enthusiastic support for the Soviet Union. But he was heavily criticized for his support of Trofim Lysenko, whose ideas were completely incompatible with mainstream genetics. This lapse of judgement on Bernal's part is difficult to understand.

The prospect of nuclear war horrified Bernal, particularly as one of his last acts before leaving government service at the end of the war had been to estimate the cost of destroying the Soviet Union — and of the Soviet Union destroying Britain. Along with Frédéric Joliot-Curie he founded the World Peace Council, which became a vehicle for Soviet propaganda but which might in the end have fulfilled its mission by exercising a restraining influence on Nikita Khrushchev during the Cuban missile crisis.

Bernal wrote profusely on science and society, and many of his revolutionary ideas on science planning are now commonplace. Towards the end of his life he became involved in planning for the Labour party, but his work

was cut short by ill health. Repeated strokes left him incapacitated and practically incommunicado — a sad end for a man whose life was based on communication.

This is a very fine (and large) book. Much more than a biography, because of Bernal's

involvement in so many sociological issues of his day, it takes the form of a social history of the first half of the twentieth century. ■

Kenneth C. Holmes is at the Max-Planck-Institut für Medizinische Forschung, Heidelberg 69120, Germany.

In search of Prometheus

Bioethics and the New Embryology: Springboards for Debate

by Scott F. Gilbert, Anna L. Tyler & Emily J. Zackin

Sinauer: 2005. 280 pp. \$14.95

James Bradley

Modern biotechnology raises some difficult questions. What is a person? What ought to be the moral status of human embryos? How should we define 'normal' and 'human'? Should we genetically engineer future generations of humans? And what about human cloning? A first step towards finding answers to difficult questions is to get the facts. Most people, even well-educated ones, do not have the facts needed to develop informed opinions on these questions. In a university cell-biology class that I recently lectured, only one of 140 students knew the source of embryonic stem cells; a multidisciplinary group of ten professors did no better.

In a remarkably far-sighted book, *The Prometheus Project* (Doubleday, 1968), physicist Gerald Feinberg wrote about humanity's need to ponder questions like these. He warned that human genetic engineering, age retardation, chemical and electrical mind modification, and artificial-intelligence technologies would force upon us irreversible decisions that should be the focus of a Prometheus (from the Greek word for 'foresight') project

that would draw on informed thought from all of Earth's peoples.

Decision time is here, but a consensus on who we are and where we are going is nowhere in sight. There has been no Prometheus project. National and state governments, private industry and even cults are going separate ways in research on embryonic stem cells, human cloning, and the creation and use of genetically modified organisms. And the strident voices of scientists, politicians and ethicists are urging us either to embrace the new biotechnologies for their potential benefits or to restrict them for fear of an undesirable future for our species.

Scott Gilbert, a respected developmental biologist and textbook author, has now entered the fray with a calm and rational voice. In *Bioethics and the New Embryology*, he and his two student co-authors provide information for students, other academics and the public about many of today's most urgent biotechnological issues.

Each of the book's seven sections explains the biology underlying a specific biotechnology and discusses the relevant ethical issues. Topics include early human development and personhood, assisted reproduction, sex selection, human cloning, stem cells, human genetic engineering, defining what is normal, genetic essentialism, and the ethics of using animals in research.



Prometheus unfound? No project to explore what we want from biotechnology has been forthcoming.

AKG-IMAGES

Ten short, interspersed essays treat specific biological, ethical and policy issues in greater depth. For example, an essay on contraception tells how a group of pharmacists that was opposed to filling prescriptions for emergency contraception had claimed that 'morning after' pills cause abortions. In fact, the evidence indicates that such pills act by preventing fertilization. This misinformation may result in preventable, unwanted pregnancies. Similarly, false claims by Vatican officials that the HIV virus can pass through intact condoms may jeopardize the lives of women in regions where AIDS is rampant. "The tragic irony," Gilbert writes, "is that being 'pro-life' regarding the creation of zygotes can make one complicit in the deaths of adult men and women." Other essays address preimplantation genetic diagnosis, umbilical-cord stem cells and eugenics. A 200-word glossary and an accompanying CD with modifiable figures and PowerPoint presentations for each chapter enhance the book's user-friendliness.

The information that Gilbert presents can largely be found in websites and other books, but nowhere else are the biological facts and the bioethical issues gathered and offered up in such a tastefully concise, beautifully illustrated, engaging and easily consumable format. For example, *Biotechnology: Demystifying the Concepts* by David Bourgaize and co-authors (Benjamin Cummings, 1999) is an excellent introductory survey of molecular genetics, genetic-engineering technologies and associated ethical and policy issues that addresses in 400 pages what Gilbert discusses in just 34. Similarly, 33 of Gilbert's pages very adequately explain the material covered in the 200-page *Human Embryonic Stem Cells* by Ann A. Kiessling and Scott Anderson (Jones & Bartlett, 2003). For cross-disciplinary perspectives on cloning, Michael C. Brannigan's *Ethical Issues in Human Cloning* (Chatham House, 2000) is very good, but so are the corresponding 16 pages in Gilbert's book. To delve deeper into any of the issues he discusses, one can download non-illustrated, book-length reports by the President's Council on Bioethics (www.bioethics.gov).

With its broad-ranging coverage of embryo-related biotechnologies, Gilbert's book makes an excellent text for high-school and university biology students and for bioethics courses. It also superbly meets the need for an accessible, accurate resource for the biotechnological knowledge needed for informed policy-making. Although lay readers may struggle with some of the science, the book is an important contribution to informed dialogue among citizens from a wide range of educational levels, professions and generations — a first step towards a Prometheus project. ■

James Bradley is in the Department of Biological Sciences, 331 Funchess Hall, Auburn University, Auburn, Alabama 36849, USA. He is currently writing a book for non-biologists called *Twenty-First Century Biotechnologies and Human Values*.

Tall tales from the deep

Singing Whales, Flying Squid and Swimming Cucumbers: The Discovery of Marine Life

by Richard Ellis

Lyons Press: 2006. 288 pp. \$24.95

Victor Smetacek

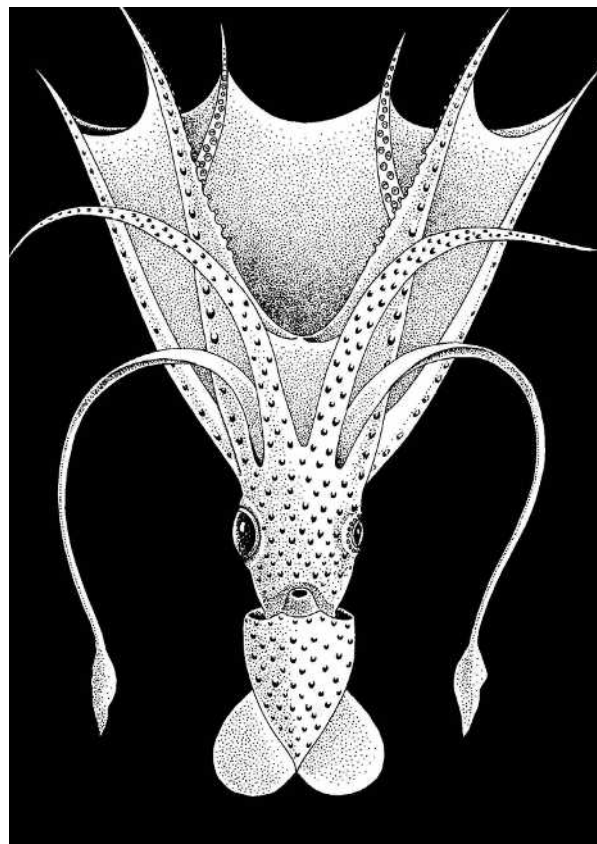
The news from the sea is depressing these days: depleted fish stocks; turtle, shark and dolphin numbers dwindling at alarming rates; Antarctic blue whales hovering on the brink of extinction; and, looming over all the wanton destruction, are the spectres of pollution, climate change and inexorable acidification. Neptune will be getting very sour. Spreading the bad news is the best one can do these days, hoping that the public at large will realize that the reports warrant red alert.

But not all is gloom and doom. Scientific understanding of how oceans and their biota function, and how they interact with the rest of the Earth system to maintain our planet in a habitable state, is growing steadily. But this endeavour of mainstream marine science is more like writing a symphony than painting a big picture. To appreciate the news, one needs to be familiar with the context. Luckily, the sea is also full (or used to be full) of eye-catching animals that appeal to the innate curiosity stimulated by large size and strangeness. Such newsworthy animals continue to be discovered, literally brought to light by the increasing number of searchlights probing the inky vastness. Spreading this type of news is also necessary, as it increases public awareness of the ocean as a fascinating and exciting habitat well worth protecting.

As suggested by its main title, Richard Ellis's book is a collection of short stories on marine animals, selected on the basis of how likely they are to fascinate the reader. The subtitle refers, with few exceptions, to discoveries made by the human eye, and hence on the periphery of mainstream research, which relies on instruments to make its observations. Ellis writes for the public in a breezy, light-hearted style, sometimes struggling to keep up the entertainment level with liberal use of superlatives. He has written many books on good and bad news from the sea, and a glance at some of the titles leaves one wondering how regular readers cope with the repetition. Perhaps the

denizens of the deep are indeed so strange that one can write several eye-catching headlines on the same topic.

Ellis is most comfortable writing about big animals and seems out of his depth in the microbial world. He says nothing about the discovery of the archaea, life-forms genetically as distant from bacteria as we are from them. They were first reported from extreme environments but are now being found thriving



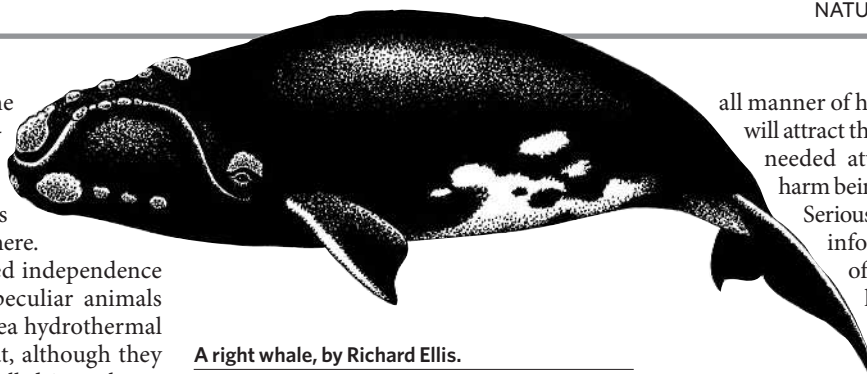
RICHARD ELLIS

Richard Ellis's drawing of an umbrella squid, from his book *Singing Whales, Flying Squid and Swimming Cucumbers*.

everywhere in oceans and lakes. The discovery of such strangeness in our midst could surely be made palatable to the public without having to tell them much about the inner life of bacteria. Instead, the single chapter on the microbial world makes news by wrongly downplaying the role of chlorophyll-driven photosynthesis in conditioning the biosphere and exaggerating that of a very different type of photosynthesis driven by rhodopsin. Both pigments capture photons to reduce carbon dioxide to organic matter, but only chlorophyll generates the herculean energy required to split hydrogen from water in order to do the job. Oxygen is the waste product here. All other types of photosynthesis get their hydrogen from energy-cheap sources such as dissolved organic matter or hydrogen sulphide. Their waste products are modified organic

molecules and sulphur. So the recent discovery of an abundance of rhodopsin-bearing bacteria in the open ocean is interesting, but not quite as Earth-shaking as portrayed here.

Similarly, the much-touted independence from solar energy of the peculiar animals that flourish around deep-sea hydrothermal vents overlooks the fact that, although they might not be fed by chlorophyll-driven photosynthesis, they are certainly breathing its waste product, oxygen. Indeed, they are as dependent on chlorophyll as all the other animals living under the Sun, down to the depths of



A right whale, by Richard Ellis.

the deepest trenches.

The book provides a healthy mixture of good and bad news from the sea. Hopefully, its portrayal of wondrous worlds inhabited by

all manner of huge and strange animals will attract the public and draw much-needed attention to the ongoing harm being inflicted on the oceans.

Serious readers looking for more information may be turned off by the overpowering hyperbole, the eclectic selection of topics, and the giddy leaps from one to another. ■

Victor Smetacek is professor of bio-oceanography at the University of Bremen and is based at the Alfred Wegener Institute for Polar and Marine Research, 27574 Bremerhaven, Germany.

RICHARD ELLIS

Seeing the light

Dan Flavin experimented with fluorescent tubes to create his art.

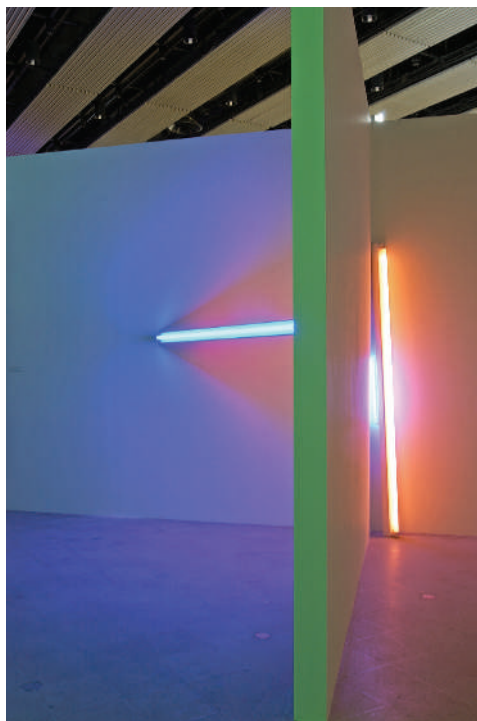
Martin Kemp

The US artist Dan Flavin (1933–96) made his lightworks predominantly using standard fluorescent tubes in five colours: red, yellow, blue, pink and green, as well as shades of white. These prosaic objects seem to provide only a basic means to achieve a limited end. Yet he aimed for nothing less than 'blank magic' — with 'blank' alluding to the plain, even banal, nature of the manufactured components he uses.

The magic lies in the extraordinary elusiveness of colour perception, particularly the effects of ambient colour, after-images, induced colours and complementary contrasts. Flavin's simple but artful arrays of glowing tubes present an extraordinary experimental field for anyone interested in the wonderful subtleties of coloured light and shadows.

Yet they are experiments without any apparent theory. He was a true 'empiric' in the ancient sense, trying things to see what happened, what worked and what didn't, learning from experience but developing no theory of causes and no system for predicting effects. He was aware of the basic theories of complementary colours (which are different for pigments and for lights) and after-images, but he never dwelt on these or other theoretical aspects of colour in his writings or interviews.

Seemingly, Flavin felt that a set theory would have led to him doing the predictable, with his art, resembling those didactic demonstrations seen in science fairs, museums and exploratoriums. Indeed, at a Flavin exhibition now showing in London, scientists Mark Lythgoe, R. Beau Lotto and Mark Miodownik have devised an informative room of demonstrations of just this kind, looking at phenomena



Light on theory: Dan Flavin's experiments with colour were empirical rather than systematic.

of colour and fluorescence.

In experimentally seeking the unknown, rather than merely repeating what is known to work, Flavin behaved like a scientist. But he stopped at his 'blank magic', never trying to propose anything more general beyond the specific experience — and this separates his work from the scientific. He serves up an elusive interpretative field, whereas Lythgoe and his colleagues provide measured demonstrations from which general principles can be adduced.

This is not to say that Flavin's work was intellectually unsophisticated. One of his works is entitled *The Nominal Three* (for Wm

of Ockham). This refers to the medieval philosopher William of Ockham, famous for his philosophical 'razor', who advocated that plurality should not be posited unnecessarily. Ockham rejected the 'universals' favoured by his contemporaries, notably Thomas Aquinas. He argued that the only reality is that of the individuality of particular things, and that universal terms such as 'man' were simply the result of the grouping of similar things by the mind. Ockham's position appealed to Flavin, given his own insistence of the reality of the experience of something, rather than the superior reality of a universal order that supplies predictive rules.

Even though Flavin was raised in an Irish Catholic family in New York, and quoted Ockham, he denied that there was any spiritual dimension in his art. However, moving round the exhibition, bathed in its soft light, with coloured shadows stalking the walls, it is difficult not to think of a church interior suffused by the glow from stained-glass windows.

Miraculous radiance was, traditionally, the manifestation of divine presence.

That Flavin's chosen medium of light is immaterial (at least to our unaided senses) lends it a mysterious, even metaphysical, dimension, beyond what he stated as his intention. His prosaic fluorescent tubes are indeed transposed into the realm of the magical. I suspect that this magic that will be familiar to any scientist who works on colour.

'Dan Flavin: A Retrospective' can be seen at the Hayward Gallery in London until 2 April, before transferring to museums in Paris and then Munich.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

www.hayward.org.uk/flavin

M. LEITH/S. FLAVIN/ARS/DACS

NEWS & VIEWS

CELL BIOLOGY

When the tail wags the dog

Scott M. Landfear

Flagella are whip-like structures that power the movement of certain cells. Analysis of a single-cell parasite, the African trypanosome, reveals that flagella are also essential for viability in this organism.

Many cells can move from one position to another, by various means. One locomotive structure that emerged early in evolution is a whip-like appendage called the flagellum that propels the cell by a cyclical beating motion. Stationary cells also use flagella and their shorter relatives, cilia, to move liquid across their surface, as in the lining of air passages for example. In nucleated, or eukaryotic, cells, the flagellum is wrapped in cell membrane and encompasses a central, flexible rod — the axoneme — that generates the beating motion. The flagella from diverse organisms have a remarkably similar morphology for the axoneme, with a cylindrical array of nine filaments called microtubules surrounding two core microtubules, which hints that the building-blocks of these structures have been highly conserved during evolution.

On page 224 of this issue, however, Broadhead *et al.*¹ reveal a surprising diversity among organisms of the roughly 300 proteins that form the microscopic bricks and mortar of the axoneme. Nonetheless, they also discovered that molecular components of the axoneme in the unicellular African trypanosome parasite are conserved from these ancient protozoa to humans. In humans, abnormal forms of these proteins are linked with debilitating diseases. Perhaps most remarkably, the authors have uncovered a more fundamental role for flagella than cellular motility. Flagella are essential for trypanosome cell division, and thus viability — an observation that may help in developing therapies against African sleeping sickness, the fatal disease caused by these parasites.

Proteomic analysis has become an increasingly powerful tool for defining the structures of subcellular bodies, termed organelles, that perform specialized functions for cells. In this approach, a particular organelle (or sub-structure thereof) is purified and the constituent

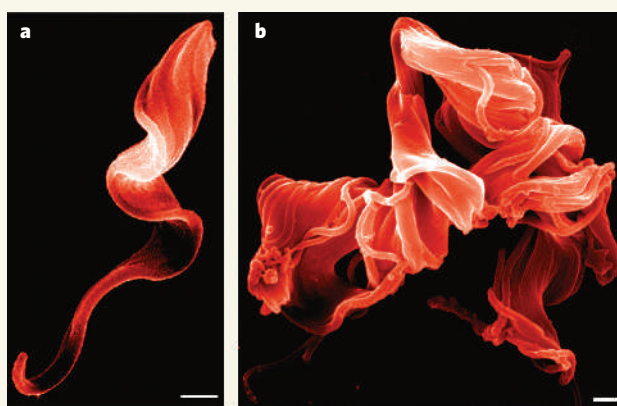


Figure 1 | Failed flail. **a**, A scanning electron micrograph of the bloodstream form of a normal African trypanosome. **b**, A parasite in which expression of one of the flagellar axoneme proteins has been destroyed by RNA interference of the corresponding messenger RNA. Broadhead *et al.*¹ found that the contorted 'monster' parasite has multiple nuclei and flagella, and cannot divide. Scale bar, 1 μ m. Another example is shown on the cover of this issue.

proteins are identified by sophisticated mass spectrometry. When Broadhead *et al.*¹ applied this proteomic analysis to isolated flagellar axonemes from the African trypanosome, they discovered that there were many proteins in the trypanosome organelle that could not be identified from the sequenced genomes of other organisms (with the exception of two other parasitic protozoa closely related to African trypanosomes). A similar analysis by these authors of recently published data on flagellar axonemes from two other unicellular eukaryotes^{2,3} revealed similarly unique cohorts of axonemal proteins in all three organisms compared. So the apparently conserved morphology of the axoneme belies a hidden molecular diversity.

All the same, there are also axonemal proteins that are conserved between trypanosomes and humans. When the authors identified the human chromosomal regions encoding all these conserved axonemal proteins, they found 34 genes that mapped to 25 loci that had been implicated previously in a plethora of genetic diseases. The characteristics of these conditions are consistent with

defects in flagella or cilia. These diseases cover a broad spectrum ranging from hydrocephalus and polycystic kidney disease to epilepsy. Although in each case a rough genetic locus had been associated with disease, the specific disease-associated genes had not been identified previously. So Broadhead and colleagues' examination of the ancient trypanosomes has done a significant service to human biology and medicine by identifying likely candidates for disease-causing genes. These results highlight the often-overlooked virtues of studying unusual or 'primitive' organisms.

But of course trypanosomes cause a serious disease in their own right. African sleeping sickness is a fatal infection that currently afflicts an estimated 300,000–500,000 people in sub-Saharan Africa⁴. How does the new work¹ help in understanding this devastating disease? This is perhaps the most intriguing part of the story. The authors examined the role of five trypanosome axonemal proteins using RNA interference, a powerful genetic method that allows the selective degradation of an individual messenger RNA, specifically reducing the encoded protein. In all five cases, the trypanosomes underwent a fatal meltdown when they could no longer make each axonemal protein. Although the parasites continued to replicate their DNA, divide their nuclei, and generate new flagella and other organelles, they were unable to divide into independent daughter cells. Instead, they formed spectacularly distorted 'monster' cells (Fig. 1) with multiple nuclei and flagella, and rapidly died.

These results were initially astonishing, but they are consistent with observations from the same laboratory that trypanosomes use their flagella to determine cell polarity during division⁵. These previous revelations hinted at what the current paper now confirms, that flagella are indispensable for telling parasites

S. GRIFFITHS/K. GULL

how to divide — especially in the stage of their life cycle when they are in the bloodstream, causing the human and animal disease. Indeed, the flagellum seems to be the tail that wags the dog. These results raise the intriguing possibility of developing drugs that act selectively against the trypanosome-specific axonemal proteins. It may be possible to identify compounds that cripple the parasite flagellum while leaving the distinct proteins of the human axoneme alone. ■

Scott M. Landfear is in the Department of Molecular Microbiology and Immunology, Oregon Health and Science University, Portland, Oregon 97239, USA.

e-mail: landfear@ohsu.edu

1. Broadhead, R. *et al.* *Nature* **440**, 224–227 (2006).
2. Pazour, G. J. *et al.* *J. Cell Biol.* **170**, 103–113 (2005).
3. Smith, J. C. *et al.* *J. Proteome Res.* **4**, 909–919 (2005).
4. www.who.int/mediacentre/factsheets/fs259/en/index.html
5. Moreira-Leite, F. F. *et al.* *Science* **294**, 610–612 (2001).

ASTROPHYSICS

Ancient blast comes to light

Enrico Ramirez-Ruiz

Light from the oldest and farthest stellar explosion yet seen was emitted when the Universe was a mere infant. It provides a close-up view of how and when stars formed, and how they affect the primordial gas around them.

A trio of contributions to this issue^{1–3} presents observations of the most distant stellar explosion ever seen: a γ -ray burst (GRB) that took place when the Universe (currently accepted age, roughly 13.7 billion years) was only about 900 million years old. Thus, for the first time, the most distant objects that can be identified spectroscopically are not just galaxies — and therefore huge agglomerates of stars, gas and dust — but also individual stars.

The hunt for cosmic structures that ignited when the Universe was in its infancy takes us to the frontier of our current observational capabilities (Fig. 1). Not too long ago, the most distant objects known were quasi-stellar objects, or quasars⁴ — glaringly luminous objects powered by gas falling into massive black holes at the centres of galaxies. Over the past few years, however, extremely sensitive surveys with space- and ground-based telescopes have allowed us to observe ordinary galaxies beyond the farthest quasars⁵.

Since then, the race has been on to find the most distant star. As most stars lead relatively unexciting lives, they remain far less luminous than galaxies, and so distant stars are generally too faint to be detected with current technology. Some heavyweight stars, however, end their lives violently and spectacularly. These stars send out bursts of radiation so luminous that they appear bright even when viewed across vast stretches of the Universe.

On 4 September 2005, such a flash of γ -rays, lasting for 80 seconds, hit NASA's purpose-built GRB-detection satellite, Swift. Swift's γ -ray monitor established the position of the burst — prosaically labelled GRB 050904 — in the constellation Pisces. As Cusumano *et al.* describe (page 164)¹, within seconds Swift turned around to direct its X-ray telescope at the region where the burst occurred, pinpointing the location of a rapidly fading source of

radiation to within a hundredth of a degree. This in turn allowed powerful optical and infrared telescopes around the world to search for the decaying 'afterglow' signal within their wavelength ranges. On page 184, Kawai *et al.*² report measurements of the afterglow of GRB 050904 at optical wavelengths. And, on page 181, Haislip *et al.*³ detail the tightly choreographed sequence of observations necessary to detect the barely visible afterglow at even longer, near-infrared wavelengths.

In astronomy, distance, time and the wavelength at which observations are made are inextricably linked. Light travels at a finite speed, and so takes a finite time to get to us. Far-off objects are thus seen as they were in the past. Cosmologists generally use the redshift z to designate distance or, equivalently, look-back time. The quantity $1+z$ is the factor by which the Universe has expanded between the time when a source emitted the light we observe and the present, and is also the factor by which the wavelength of light reaching us has been stretched owing to the expansion of the Universe.

So by measuring the wavelengths of absorption lines in the spectrum of the optical afterglow of GRB 050904, and comparing these with wavelengths characteristic of the heavier elements such as carbon, sulphur and silicon that are found in stars, the redshift of GRB 050904 was found² to be $z = 6.295$. It is thus the most distant stellar explosion ever witnessed, and occurred when the Universe was about 7% of its current age. Setting a new record for the farthest stellar explosion might seem noteworthy enough, but distance alone does not make GRB 050904 interesting. Its attraction lies in what it can reveal about star formation early in the history of the Universe.

Over the past decade, astronomers have been particularly successful in tracing how

early stellar systems changed and coalesced to become the 'respectable' elliptical, spiral and irregular galaxies seen in the present-day Universe. These early galaxies are far less luminous than quasars, but are still (as we see them) forming new stars, and so emitting copious amounts of light. Among these, the galaxy hosting the youngest-known quasar⁴ gives the earliest example of a star-forming region⁶. The presence in this galaxy of a large amount of mass in the form of interstellar molecules and dust is proof not only that substantial star formation was taking place only 800 million years after the Big Bang, but also that the process of chemical enrichment — in which heavier elements synthesized by previous generations of stars are incorporated into the galaxy — was well advanced.

Star formation in these early epochs was therefore probably not too different from that taking place now in some regions of the local Universe. It is in itself no surprise that a GRB can occur at these early times: GRBs are thought to be produced by the deaths of a rare breed of massive stars⁷, and, the more massive a star, the shorter its lifetime. Stars slightly more massive than our Sun have lifetimes that are shorter than the age of the Universe at $z = 6.295$, and a star 20 times as massive can keep going for only a thousandth as long.

So far, however, GRB 050904 is the only example of a stellar explosion that occurred at a time in the evolution of the Universe before 'reionization', the point when the hydrogen between galaxies was completely ionized by ultraviolet radiation. Neutral hydrogen in the Universe at $z = 6.295$ would have been a very effective absorber of any photon from GRB 050904 emitted at a wavelength shorter than the hydrogen 'Lyman α ' line at 121.6 nm. Interestingly, such a damped Lyman α absorption profile of neutral hydrogen is not easily discernible in the optical afterglow spectrum² of GRB 050904. Instead, one sees flux absorbed by the presence of a dense screen of metal-enriched gas in the immediate surroundings, presumably ejected by the massive stellar progenitor prior to collapse. This might make GRBs less suited as clean markers of when, exactly, the reionization epoch occurred⁸. More optimistically, the detection of metal absorption lines in the afterglow spectrum from the gas around the burst site offers an unusual opportunity to study the physical conditions in the interstellar medium at these early epochs.

GRB observations will certainly pave the way for the discovery of more primitive objects than the galaxies presently known at $z = 6.5$. Distant galaxies are themselves difficult to study because they appear faint and small. GRBs will serve not only as signposts to such galaxies, but could be used to study the gradual build-up of heavy elements in them to determine the conversion history of primordial gas into stars. Although the light from GRB 050904 is likely to have come from a

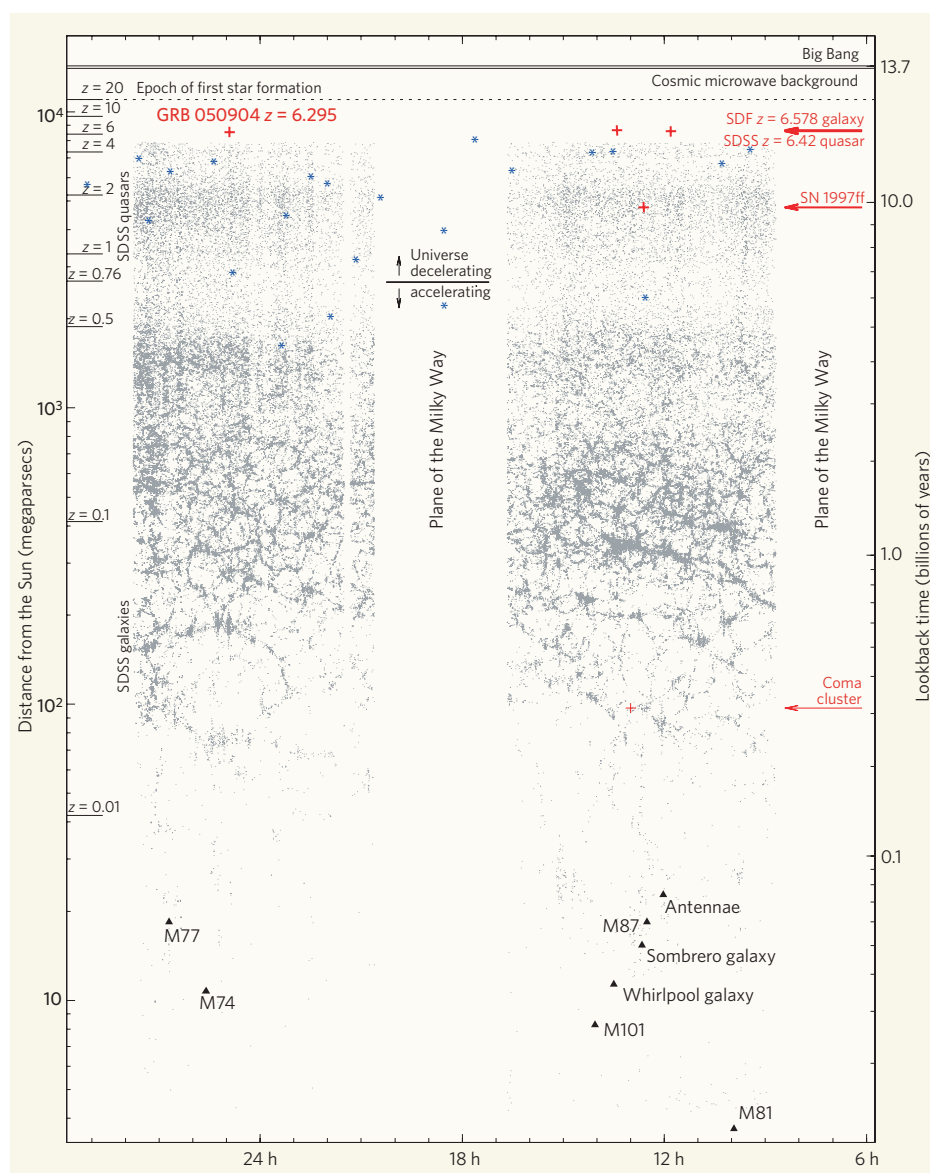


Figure 1 | The Universe as we know it. A 360° vista showing the entire sky, with visible structures stretching back in distance, time and redshift. (The immediate light and dust from our own Galaxy, the Milky Way, obscures all light from outside sources in its plane.) The most distant light we observe comes from the radiation left over from the Big Bang: the cosmic microwave background. As we descend the chart, we find the most distant objects known, followed by a web of Sloan Digital Sky Survey (SDSS) quasars and galaxies. Closer to home, we start to see a collection of familiar 'near' galaxies (triangles). At $z = 0.7554$ (in the favoured cosmological models), the Universe undergoes a transition from a slowing expansion to an accelerating one. Objects with $z < 0.7554$ are observed at an epoch when the Universe's expansion is accelerating, whereas objects farther away are observed at an epoch when the Universe's expansion is decelerating. Also marked are all Swift GRBs with known distances (blue stars); SN 1997ff, the most distant type Ia supernova at $z = 1.7$; and the archetypal large galaxy cluster, the Coma cluster. The most distant GRB, reported in this issue¹⁻³, is at redshift $z = 6.295$, comparable to the most distant galaxies found by the Subaru Deep Field (SDF) survey, and the most distant SDSS quasar. (Courtesy of J. R. Gott and M. Juric¹⁰.)

later star that was already metal-enriched, bursts arising from the very first generation of stars⁹ (those comprising only light nuclei such as hydrogen and helium produced in the Big Bang) may be detectable. And if bursts can be detected from a time before galaxies had gravitationally assembled, they might even provide a glimpse into the pregalactic phase of the Universe. ■

Enrico Ramirez-Ruiz is at the Institute for Advanced Study, Einstein Drive,

Princeton, New Jersey 08540, USA.
e-mail: enrico@ias.edu

1. Cusumano, G. *et al. Nature* **440**, 164 (2006).
2. Kawai, N. *et al. Nature* **440**, 184-186 (2006).
3. Haislip, J. B. *et al. Nature* **440**, 181-183 (2006).
4. Fan, X. *et al. Astron. J.* **125**, 1649-1659 (2003).
5. Kodaira, K. *et al. Publ. Astron. Soc. Jpn* **55**, L17-L21 (2003).
6. Walter, F. *et al. Nature* **424**, 406-408 (2003).
7. Hjorth, J. *et al. Nature* **423**, 847-850 (2003).
8. Barkana, R. & Loeb, A. *Astrophys. J.* **601**, 64-77 (2004).
9. Abel, T. *et al. Science* **295**, 93-98 (2002).
10. Gott, J. R. *et al. Astrophys. J.* **624**, 463-484 (2005).



50 YEARS AGO

Ageing in Industry by F. Le Gros Clark and Agnes C. Dunne — The only people who seem to have solved the problem of their senescent members seem to have been the mythical Hyperboreans whose aged, when they wished to retire from life, performed a "joyful suicide" by decking themselves with garlands and precipitating themselves from a rock into the sea. Non-mythical peoples have found various other solutions. The Tiberians used to hang their old on gibbets, the Hircanians and Bactrians cast them, while still living, to the dogs, and the Scythians buried them alive.

In Great Britain, such drastic measures have never been contemplated, although the plight of the aged has often given cause for much distress. During the past two or three decades we have become increasingly conscious of the wider repercussions of our ageing population and the resulting social and economic disadvantages.

From *Nature* 10 March 1956.

100 YEARS AGO

Geschichte der biologischen Theorien, seit dem Ende des siebzehnten Jahrhunderts by Dr. Em. Radl — Although biology is now permeated by the evolution idea, and has continually before it the ideal of giving a genetic description of the present phase of the animate world, there is some reason to fear, as Dr. Radl indicates, a growing apathy towards the study of the evolution of the science itself. Whether it be that many workers share Nietzsche's view that the study of history paralyses the intelligence, or that they feel it their primary business to make history, not to read it, or that they regard historical inquiries as the philosopher's task... The modern biologist, intent on new discoveries, has no use for Aristotle, Descartes and Leibnitz, but their influence may be upon him none the less... even the most modern system of biology is, like our own body, a museum of relics.

From *Nature* 8 March 1906.

50 & 100 YEARS AGO

BIOLOGICAL CHEMISTRY

Catalytic competition for cells

Virginia W. Cornish

Ways of evolving proteins, and assessing the vast numbers of variants needed to identify those with novel enzymatic activity, are themselves evolving. Oil droplets containing basic cell machinery provide a promising approach.

The cell has the enviable ability to evolve through mutation of its hereditary DNA code. When we first learned how to mutate DNA in the test tube, and so manipulate the amino-acid sequence of a protein, we quickly learned how difficult it is to rationally alter protein function with just a few amino-acid changes¹. However, modern DNA technology makes it possible to generate not one or two but 10^{10} or more protein variants with an altered amino-acid sequence, and we can now carry out 'directed evolution' — of, for example, surrogates of green fluorescent protein that range in colour from cyan to red². But what will it take to compete with the evolutionary power of the cell to create and identify even more dramatic changes in function?

As described in back-to-back papers published in *Chemistry & Biology*, groups led by Dan Tawfik³ and Andrew Griffiths⁴ attempt to extend the directed evolution of enzyme catalysis to chemistry beyond that naturally carried out in the cell. Proteins with new functions presumably evolve in the cell through the accumulation of mutations in genomic DNA generated by random genetic drift, followed by selective amplification of cells with the fittest variants when some selective pressure is applied. Directed evolution seeks to recapitulate this process on an experimentally accessible timescale by selectively introducing mutations into the DNA encoding the protein of interest at a high rate, and then picking the handful of protein variants that have acquired the desired new function⁵.

Synthesizing 10^{10} protein variants at the DNA level is, in fact, easy. Variations of the polymerase chain reaction, a technique for selectively amplifying a segment of DNA, make it possible not only to mutate select amino acids in the active site of a protein, but also to replace a whole loop in a protein or mimic natural recombination by swapping whole segments from a related protein sequence. (Note, however, that 10^{10} sequence variants is a tiny number compared with all the possible sequence variants for even a 200-amino-acid protein composed of 20 different amino-acid building-blocks.)

But with such large numbers, identifying the handful of proteins with the desired new function is very hard. The cell offers an elegant solution to this problem of finding the needle in the haystack. It acts as a self-replicating compartment that links a mutable and amplifiable DNA code to the catalytic

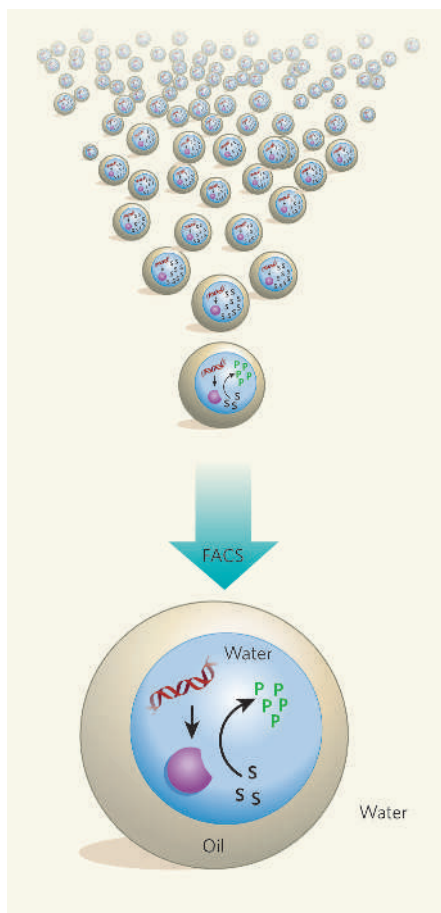


Figure 1 | Finding the needle in the haystack^{3,4}. Water-in-oil-in-water emulsions of the cell's protein-synthesis machinery allow fluorescence-activated 'cell' sorting (FACS) of individual droplets containing not only a protein and its unique DNA sequence, but also fluorogenic reporters for different chemical transformations. Individual droplets that contain enzyme variants (purple) with increased catalytic activity can be sorted at a rate of about 10^7 per hour based on the number of fluorescent product molecules (P) synthesized from the substrate (S). The DNA (red) from active droplets can then be recovered and amplified.

translated from that DNA, and to the myriad functions that determine the fitness of the cell. Some methods (phage display, for example) can directly or indirectly link each protein physically to its unique DNA sequence, even for 10^{10} variants, and have been exploited for the *de novo* evolution of binding proteins⁶. But they do not lend themselves readily to high-throughput assays for enzyme catalysis.

Traditional enzyme assays can be carried out one-by-one in microtitre plates using automation techniques. But signal-to-noise issues limit these assays to smaller numbers of protein variants in practice. For the vast range of chemical transformations not carried out in the cell, we do not have the assays to sort through the large number of protein variants, and the directed evolution of *de novo* catalysts eludes us.

Tawfik, Griffiths and their co-workers have confronted this problem by developing an 'in vitro compartmentalization' (IVC) technology that basically strips the cell's machinery for transcribing and translating DNA to RNA to protein, and reconstitutes it in water-in-oil droplets that have about the same volume as a bacterial cell. This approach provides one solution for linking each unique protein variant to its DNA sequence, because statistically it is easy to create 10^{10} water-in-oil droplets each containing a unique DNA sequence encoding a unique protein variant.

Tawfik and Griffiths have already successfully used their IVC technology for test-tube evolution of proteins, but largely for enzymes involved in modifying DNA. In the new papers^{3,4} they go further, providing not only a link between the DNA and the protein it encodes, but also a functional assay that can handle large numbers of variants.

Both groups show that they can make water-in-oil-in-water emulsions that allow the encapsulation of fluorogenic indicator dyes used to detect enzyme catalysis in more traditional formats. They then submit some 10^7 droplets, each containing a unique protein variant, to a technique known as fluorescence-activated 'cell' sorting (FACS; Fig. 1). Using FACS on these water-in-oil-in-water emulsions, they can carry out test-tube evolution to increase the catalytic activity of a known protein.

Tawfik and co-workers³ leave behind the IVC technology and literally encapsulate a bacterial cell. They show that this increases the concentration of enzyme that can be produced in each individual droplet to about 10^5 molecules. From just one round of mutation and FACS screening, they then isolate a variant of the natural enzyme paraoxonase with a 100-fold increase in hydrolytic activity from 10^6 different variants. Griffiths and co-workers⁴ carry out FACS with cell-free IVC droplets, which synthesize about 100 copies of protein per droplet. Returning to a classic experiment in directed evolution, they evolve Ebg, a protein of unknown function made by the bacterium *Escherichia coli*, into a β -galactosidase enzyme using a fluorogenic β -galactosidase substrate. With multiple rounds of mutation and screening, they isolate several Ebg variants showing a more than 300-fold increase in β -galactosidase activity compared with Ebg.

So what is the best way to compete with the evolutionary power of the cell? Tawfik and Griffiths strip the cell of some of its basic machinery, and, by analogy to the cell,

compartmentalize this machinery in a water-in-oil droplet. Their IVC system preserves the cell's means of linking DNA to protein, and then adds on *in vitro* chemistry to create the evolutionary pressure. At the other extremes are completely synthetic encoded systems⁷⁻⁹, or solutions that seek to expand the chemistry carried out by the cell¹⁰. The advantage of a completely synthetic system may be that it can go beyond the chemistry that can be synthesized or tolerated by the cell, although the range of chemistry naturally carried out by the cell is awfully impressive.

The field of directed evolution is in a vibrant phase^{11,12}. Beyond that, this tinkering with cells

will provide useful technologies for genomics and biomedical research, and will inspire thinking about what might be synthesized to recapitulate the functions of the cell and how the cell might be co-opted for new functions^{13,14}. Virginia W. Cornish is in the Department of Chemistry, Columbia University, Havemeyer Hall, MC 3111, 3000 Broadway, New York, New York 10027, USA.

e-mail: vc114@columbia.edu

1. Knowles, J. R. *Science* **236**, 1252-1258 (1987).
2. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. *Nature Meth.* **2**, 905-909 (2005).
3. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D. S. *Chem. Biol.* **12**, 1281-1289 (2005).

4. Mastrobattista, E. *et al. Chem. Biol.* **12**, 1291-1300 (2005).
5. Arnold, F. H. *Nature* **409**, 253-257 (2001).
6. Binz, H. K., Amstutz, P. & Pluckthun, A. *Nature Biotechnol.* **23**, 1257-1268 (2005).
7. Sculimbrene, B. R. & Miller, S. J. *J. Am. Chem. Soc.* **123**, 10125-10126 (2001).
8. Gartner, Z. J. *et al. Science* **305**, 1601-1605 (2004).
9. Halpin, D. R. & Harbury, P. B. *PLoS Biol.* **2**, e174 (2004).
10. Lin, H., Tao, H. & Cornish, V. W. *J. Am. Chem. Soc.* **126**, 15051-15059 (2004).
11. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. *Nature advance online publication* doi:10.1038/nature04607 (2006).
12. Park, H. S. *et al. Science* **311**, 535-538 (2006).
13. Szostak, J. W., Bartel, D. P. & Luisi, P. L. *Nature* **409**, 387-390 (2001).
14. Benner, S. A. & Sismour, M. *Nature Rev. Genet.* **6**, 533-543 (2005).

MOLECULAR PHYSICS

Recombination cool and fast

Benjamin J. McCall

Molecular physicists and astrophysicists alike would like to know how fast the H_3^+ molecular ion recombines with electrons. Fast, seems to be the answer — with an awkward consequence for the astrophysicists.

Every schoolchild knows that, like opposing poles of a magnet, opposite charges attract. But what happens when charged bodies are small enough that the rules of quantum mechanics come to bear, for instance when an electron and a positively charged molecule attract? Here, the situation is more complicated: even the reaction between an electron and the simplest polyatomic molecule, H_3^+ (which can be thought of as a hydrogen molecule, H_2 , with an extra proton, H^+) has puzzled both theorists and experimentalists for decades. In a contribution to *Physical Review Letters*, Kreckel *et al.*¹ describe an ingenious experiment that provides further elucidation of the speed of this fundamental reaction.

When an electron approaches a singly charged positive ion (call it X^+), both bodies experience an attraction that accelerates them and causes them to collide. They can recombine to form neutral X, provided that the extra kinetic energy that they have gained by being accelerated can somehow be removed. For macroscopic objects, this is not generally a problem: friction dissipates the energy. At the quantum-mechanical level, however, this cannot happen. If X^+ is the ion of a single atom, energy can be lost only by emitting a photon, a slow process that seldom happens during the short time a collision takes. In most such collisions, the ion and electron fly away from each other again. If X^+ is a molecular ion, however, there is a much more efficient option: the molecule can break apart following recombination with the electron, and the resulting neutral fragments can carry away kinetic energy. This is the process known as dissociative recombination.

The H_3^+ ion assumes an important role in

astrophysics as the first link in a chain of chemical reactions in interstellar clouds through which most of the molecules found in interstellar space form (Fig. 1). Interstellar clouds were recently seen to contain much more H_3^+ than expected², bringing the persistent enigma of its recombination rate back to the fore. Exactly how the dissociative recombination of H_3^+ works was explained

theoretically only recently^{3,4}, and, starting in 1973, many experimental measurements have yielded drastically differing values for the rate at which it occurs⁵.

H_3^+ is produced in ionized gases known as plasmas. Different plasma conditions will lead to different degrees of vibrational and rotational excitation of the H_3^+ ions, perhaps accounting for some of the variation in the experimental recombination rates. If the rate of recombination were lower than assumed, especially at the lower temperatures of interstellar space, the overabundance of H_3^+ in diffuse clouds could be easily explained. But without accurate values for the rate, further progress in understanding the mystery of H_3^+ abundance is impossible.

Hence the efforts to understand dissociative recombination in terrestrial laboratories. In order to best simulate the conditions in

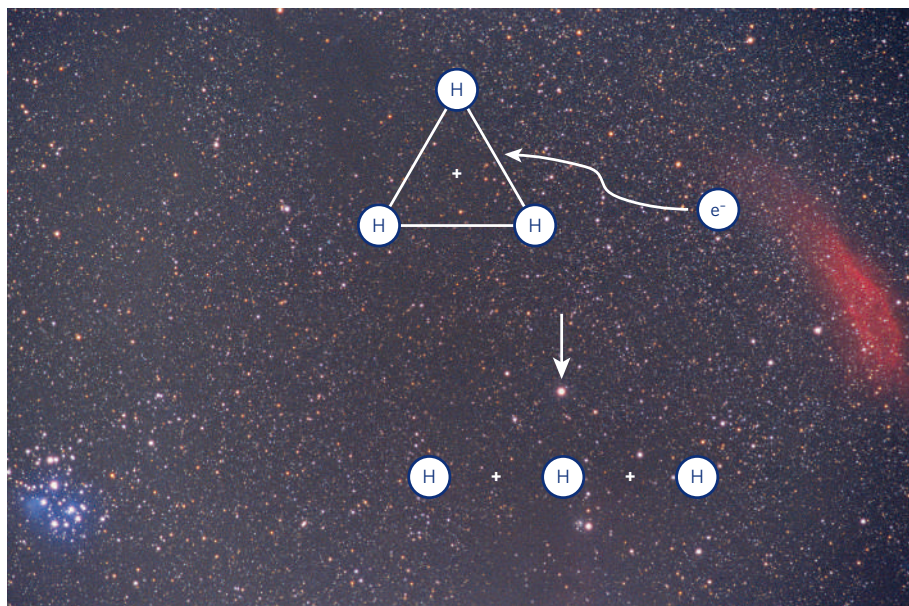


Figure 1 | Cosmic mystery. When an electron and a positively charged H_3^+ ion meet, three separate neutral hydrogen atoms are formed in a process known as dissociative recombination. The image is overlaid on a wide-field image of the Perseus region: the bright stars at bottom left are the Pleiades, or Seven Sisters, and the red region to the right is the California nebula. The bright star at the arrow's head is ζ Persei, where H_3^+ has been observed in unexpectedly high abundance². The results of Kreckel *et al.*¹ imply this cannot be due to slow dissociative recombination. The most likely explanation is instead an enhanced rate of ionization by cosmic rays.

interstellar space, the ions in the laboratory must be cooled to a state of minimum rotation and vibration. Vibrational cooling was first achieved using the CRYRING storage ring in Stockholm, Sweden, by injecting mass-selected ions into an accelerator, and allowing them to relax to their vibrational ground state⁶. Achieving a low rotational temperature is more difficult: H_3^+ cannot cool rotationally by emitting radiation⁷, so the ions must be prepared in a rotationally cold state before they are injected into the storage ring. At CRYRING, this is done by sparking a discharge to ionize a jet of hydrogen gas as it expands through an opened valve into the vacuum of the storage ring and cools⁸. Once the H_3^+ ions are in both their vibrational and rotational ground states, an electron beam travelling at a well-defined velocity is merged with the ion beam, and dissociative recombination occurs. The neutral fragments produced are counted as a function of the relative velocity of the two beams, yielding the 'spectrum' of the rate of recombination as a function of the collision energy.

Kreckel and colleagues' experiments¹, performed at the Test Storage Ring (TSR) in Heidelberg, Germany, follow a similar scheme, but make two improvements over the previous studies. First, they produced their electron beam with a newly developed cryogenic photocathode, allowing more precise control of the ion–electron collision energy and so higher resolution in the dissociative-recombination spectrum. Second, they used a new type of ion source, called a radiofrequency multipole ion trap. In such a trap, H_3^+ ions are stored before injection into the ring at low temperature in the presence of helium gas. The large number of collisions of the ions with the helium means that the rotational energy of the ions can be transferred to the helium, thus ensuring that the ions are rotationally cold. (In contrast, with the expanding jet source used at CRYRING, a small fraction of ions may remain rotationally warm.)

The result of the TSR experiment is in excellent agreement with the CRYRING results, and, thanks to its higher resolution, reveals the fine detail of the spectrum more clearly. It confirms that the rate of dissociative recombination is fast under cool, interstellar conditions. Improved theoretical calculations^{3,4} also yield a rate that agrees well with both experiments. Some minor discrepancies remain, but the general concord implies that the long-standing enigma of the rate of H_3^+ recombination might finally be resolved. If so, the onus is back on the astrophysicists: how can the large observed abundance of H_3^+ in diffuse clouds be explained if recombination is so fast? One likely solution would seem to be enhanced production of H_3^+ through ionization by cosmic rays².

Kreckel and colleagues' results¹ do contain an intriguing twist. At the low temperatures of their measurements (and of the interstellar medium), H_3^+ exists almost entirely in its two lowest rotational states, which have a total

nuclear spin of 3/2 (ortho- H_3^+) and 1/2 (para- H_3^+). But by using para- H_2 in their ion source, and thus enhancing the ratio of para- to ortho- H_3^+ , the authors saw a marked difference in the rate of dissociative recombination at low energies. Unfortunately, they were not able to measure the degree of the para- H_3^+ enhancement, and because of the nature of their ion source, it is probably not very large. Future experiments with pure para- H_3^+ would be highly desirable to elucidate the difference in the rate of recombination between the two states. That would indeed represent the first dissociative recombination measurement of a single quantum state. ■

Benjamin J. McCall is in the Departments of Chemistry and Astronomy, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA.

e-mail: bjmcCall@uiuc.edu

1. Kreckel, H. *et al.* *Phys. Rev. Lett.* **95**, 263201 (2005).
2. McCall, B. J. *et al.* *Nature* **422**, 500–502 (2003).
3. Kokouline, V., Greene, C. H. & Esry, B. D. *Nature* **412**, 891–894 (2001).
4. Kokouline, V. & Greene, C. H. *Phys. Rev. A* **68**, 012703 (2003).
5. Larsson, M. *Phil. Trans. R. Soc. Lond. A* **358**, 2433–2444 (2000).
6. Larsson, M. *et al.* *Phys. Rev. Lett.* **70**, 430–433 (1993).
7. Kreckel, H. *et al.* *New J. Phys.* **6**, 151 (2004).
8. McCall, B. J. *et al.* *Phys. Rev. A* **70**, 052716 (2004).

NEUROBIOLOGY

How fast can you go?

Laura N. Borodinsky

Rhythmic activities such as walking need tight coordination. In mice, pace is tweaked by a specific set of spinal-cord neurons that, surprisingly, make the animals walk faster by inhibiting the underlying circuit.

Watch your step — walking may seem simple, but is actually quite a complex task. As with other rhythmic motor behaviours (breathing or swallowing, say), locomotion relies on a finely tuned neuronal network that is headquartered in the spinal cord^{1,2}. The ensemble of spinal neurons that generates a coordinated rhythmic activity is known as a central pattern generator. The rhythm and periodicity of this network determines movement features such as the alternation between left and right, or the speed of walking, jumping or swimming³. Understanding how this circuit operates and the specific roles of the different neurons that participate in it is difficult, but Gosgnach and colleagues⁴ have taken up the challenge. On page 215 of this issue, they report that the activity of a group of spinal-cord neurons controls the speed of locomotor behaviour in the mouse.

Neuronal circuits are formed by a network of interconnected excitatory and inhibitory neurons. In a very simplistic model, the former group switches the circuit on and the latter turns it off. Gosgnach *et al.* studied the role of a subclass of inhibitory spinal neurons known as V1 neurons, which are thought to be part of the central pattern generator. The researchers used detailed information about the gene-regulatory factors that underlie the development and specialization of these neurons, to generate mutant mice in which V1 neurons were either eliminated or silenced acutely during the experiment, and examined the changes in locomotor activity.

Counterintuitively, they found that removing the inhibition caused by V1 neurons slows the speed of the locomotor rhythm. Mutant

mice lacking V1 neurons are unable to walk fast, but they can maintain normal motor behaviour at a slower pace. The authors demonstrate that this is because motor neurons connecting the spinal cord to the muscle are not sufficiently inhibited, because connections from the V1 neurons are missing in the mutant animals. The activity of motor neurons needs to be interleaved with precise periods of silence to generate a faster pace. These results underscore the value of inhibition in the nervous system: the delicate balance and fine-tuning of neuronal activity set by inhibitory connections is not only important in quieting down the system, but can also change core features of nervous-system function.

Even though Gosgnach and colleagues' mice slow down, other parameters of their locomotor activity remain intact, such as the alternation of left and right limbs necessary for coordinated stepping. As the authors showed previously⁵, a different class of spinal neurons (V0) is responsible for left–right coordination. In mutant mice lacking V0 neurons, the left and right motor neurons fire at the same time, rather than alternating. These are significant insights into how the work is distributed among the vast collection of spinal neurons that make up the central pattern generator.

How are the circuits of the central pattern generator established? When exploring the formation of circuits, scientists have focused on two principal alternative theories. One of these proposes that a genetically driven programme predetermines the identity of the neurons that participate in a given circuit and dictates how and between which of them connections are made⁶. The second proposes that

early in development, neurons exhibit spontaneous electrical activity that affects the process of neuronal specialization and the establishment of appropriate connections^{7,8}. Changes in activity during development can indeed lead to resetting of the intrinsic excitability of neurons as well as reconfiguration of connections⁹. In particular, formation of motor central pattern generators requires early spontaneous electrical activity, because when this activity is disrupted, motor neuron fibres fail to follow their normal trajectories¹⁰.

However, Gosgnach *et al.*⁴ find that the activity of V1 neurons does not seem to be involved in the formation of the central pattern generator. Their results show that loss of V1 neurons early in development is not compensated for by any reconfiguration of the circuit — the mutant mice that chronically lost V1 neurons at early stages exhibit the same slow locomotor rhythm as that observed when neurons are acutely silenced after the circuit is formed.

Nevertheless, studies of rewiring during development are complex; they require a rather dynamic approach because exclusive examination of the final end point may not reveal intermediate remodelling processes that are crucial to the resulting network. In this regard, it would be interesting to test what happens to the circuit and locomotor activity if neurons are silenced at different times during development. Formation of circuits is likely to depend on both genetically driven and activity-shaped processes. We need to keep our minds open if we are to understand the interplay of these driving forces.

Developmental biologists and physiologists have tended to approach the problem of circuit formation and function from very different angles, without much dialogue between them. But genetic tools are now available to reconcile the results from these two disciplines and establish a fruitful interaction. Information about signatures of gene-regulatory factors that determine the fates of neuronal populations can now serve as tools with which to elucidate the functions of a given set of neurons. The work presented by Gosgnach *et al.* sets up a firm bridge across the river that has so far divided developmental biology and physiology.

Laura N. Borodinsky is in the Neurobiology Section, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0357, USA. e-mail: lborodin@biomail.ucsd.edu

- Sherrington, C. J. *Physiol. (Lond.)* **47**, 196–214 (1913).
- Hamburger, V., Wenger, E. & Oppenheim, R. J. *Exp. Zool.* **162**, 133–160 (1966).
- Brown, T. *Proc. R. Soc. Lond. B* **84**, 308–319 (1911).
- Gosgnach, S. *et al. Nature* **440**, 215–219 (2006).
- Lanuza, G. M., Gosgnach, S., Pierani, A., Jessell, T. M. & Goulding, M. *Neuron* **42**, 375–386 (2004).
- Tanabe, Y. & Jessell, T. M. *Science* **274**, 1115–1123 (1996).
- Gu, X. & Spitzer, N. C. *Nature* **375**, 784–787 (1995).
- Catalano, S. M. & Shatz, C. J. *Science* **281**, 559–562 (1998).
- Turrigiano, G. G. & Nelson, S. B. *Nature Rev. Neurosci.* **5**, 97–107 (2004).
- Hanson, M. G. & Landmesser, L. T. *Neuron* **43**, 687–701 (2004).

BIOENGINEERING

Diagnosis on disc

Frances S. Ligler and Jeffrey S. Erickson

Highly complex immunoassays that identify and quantify many different antigens simultaneously need high-resolution imaging capability. A simple, low-cost technique could be music to our ears.

The capabilities of immunoassays — tests that use the binding of ‘capture’ antibodies to antigens in order to identify the latter — have been advancing in leaps and bounds over the past three decades. Methods of enzyme amplification have increased assay sensitivity. Diode lasers have reduced the size and cost of instrumentation. Microfluidics has enabled both the analysis of very small sample volumes and the parallel processing of multiple samples. And microarrays of capture antibodies, or other recognition molecules, attached to surfaces have made possible the simultaneous testing of a sample for large numbers of target molecules.

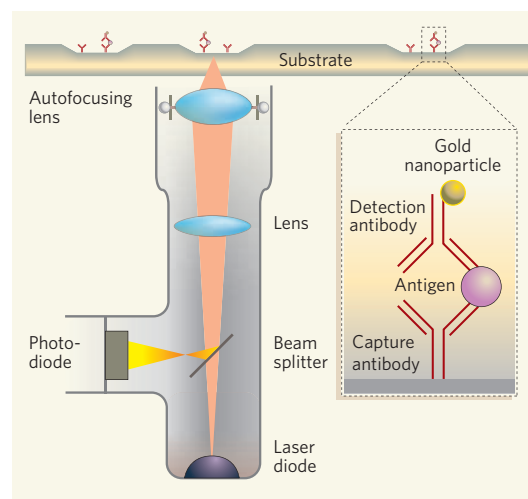
This last point harbours a problem. As microarray elements become smaller and smaller for larger and larger numbers of simultaneous tests, the necessary imaging resolution cannot readily be achieved with the standard, off-the-shelf CMOS or CCD technologies. Confocal scanners using microscope lenses can provide the required resolution, but their cost, size and the geometrical alignment of their optics pose other practical problems. Writing in *Angewandte Chemie International Edition*, Sebastian A. Lange, Günter Roth and colleagues find a way out of this seeming impasse¹. They demonstrate an immunoassay readout of high efficiency and sensitivity — using, in a slightly modified form, a pick-up head of a compact-disc player (Fig. 1).

The authors’ system is both elegant and simple. They used a so-called sandwich immunoassay, in which the antigen to be measured is bound between two different antibodies. First, a capture antibody is stamped in a 25-micrometre square pattern onto a solid

substrate; the desired antigen binds to this antibody. Second, a detector antibody binds to this antigen to make it visible to the CD pick-up head. Here, the authors cleverly employed an antibody carrying a gold nanoparticle. This gold nanoparticle catalyses the deposition of silver grains onto the substrate to which the capture antibody, antigen and detector antibody are all now attached. The reflections from the silver grains could be read by the CD pick-up head with a resolution of around 50 nanometres — considerably smaller than the diameter of most of the silver nanoparticles, which is of the order of several hundred nanometres. Using their sandwich system, the authors demonstrated that they could detect antigen in serum at concentrations from 1 microgram per millilitre down to 100 picograms per millilitre.

The authors also performed a carefully controlled study in which they spaced antigens, each attached to a capture molecule, well apart from each other on the substrate¹. They suggest that, by correlating the density of the silver precipitate with the coverage of the antigen, it might be possible to detect single molecules with their system. But measuring a single molecule carefully tied to a uniform surface is a far cry from pulling a single molecule out of a solution, capturing it at a surface and measuring it against a background signal generated from any of several variable sources. The potential for using the CD pick-up head with silver staining for single-molecule detection is therefore less convincing than the suggestion that the technology could be widely useful for low-cost, high-sensitivity readout

Figure 1 | Molecular music. A conventional compact-disc pick-up reader works by focusing laser light onto the surface of the CD. As the CD rotates above the reader, information encoded as pits along a spiral track on its metal-coated surface can be read by means of light reflected back through a lens onto a photodiode. Lange and colleagues’ immunoassay system¹ works in exactly the same way, but with a substrate covered with a regularly spaced array of capture antibodies taking the place of the CD. When an antigen binds to an antibody, its presence is signalled by a second, gold-containing detector antibody that catalyses the deposition of reflective silver particles onto the substrate.



systems for many different types of assay.

Significant issues remain to be addressed. If the CD pick-up head is to be used to detect many different molecules simultaneously, multiple capture molecules will have to be immobilized on surfaces suitable for use with the CD pick-up heads. This is almost sure to require a more flexible technology than the stamping method used by Lange *et al.* Optical alignment issues will also need to be addressed. In Lange and colleagues' study, the CD pick-up head is only one part of the optics: it is combined with a microscope stage for coarse adjustment of the magnification, as well as a lateral translation stage to adjust the position of the line of sight over the substrate. Although it may be possible to use built-in hardware in a commercially available CD player to perform these functions, this has yet to be demonstrated.

One noteworthy advantage of the CD-pick-up approach, alongside small size, low cost and high resolution, is not emphasized by the authors. Although an assay signal has never actually been generated using a CD as a sensing surface, biochemical manipulations have already been performed directly on a CD.

Externally readable fluorescence signals have even been generated in channels within the disc²⁻⁵ — the challenge here being the use of centrifugal force to move the fluids through the processing steps on the surface of the CD to the readout position. The combination of such fluidic approaches with *in situ* signal generation as demonstrated by Lange *et al.*¹ could potentially lead to a sea change in medical diagnostics. Imagine in the future buying a 'respiratory pathogen CD' from the local pharmacy when you catch a cold, inserting your self-test swab and placing it in your portable player to find out whether you should take antibiotics or stick to the chicken soup. ■ Frances S. Ligler and Jeffrey S. Erickson are in the Center for Bio/Molecular Science and Engineering, Naval Research Laboratory, Code 6900, Washington DC 20375-5348, USA. e-mails: fligler@cbmse.nrl.navy.mil; jerickson@cbmse.nrl.navy.mil

1. Lange, S. A. *et al.* *Angew. Chem. Int. Edn* **45**, 270-273 (2006).
2. Rothert, A. *et al.* *Anal. Biochem.* **342**, 11-19 (2005).
3. Gustafsson, M., Hirschberg, D., Palmberg, C., Jorvall, H. & Bergman, T. *Anal. Chem.* **76**, 345-350 (2004).
4. Tibbe, A. G. J. *et al.* *Cytometry* **47**, 173-182 (2002).
5. Lai, S. *et al.* *Anal. Chem.* **76**, 1832-1837 (2004).

PHYSICAL CHEMISTRY

Stressed molecules break down

Steve Granick and Sung Chul Bae

Tough carbon-carbon bonds can snap in certain large molecules just because the two sides of the molecule cannot agree on which way to go during adsorption. Heresy? The view through the microscope suggests otherwise.

On page 191 of this issue¹, Sheiko and colleagues show that the mechanical deformation induced merely by the adhesion of a complex molecule to a surface can trigger the break-up of that molecule. They thus provide convincing support for the seemingly heretical notion that the commonplace and unremarkable process of adsorption to a surface can bring about what otherwise occurs only with the greatest effort: the rupture of the strong, covalent carbon-carbon bond.

The system designed by the authors¹ is elegant in its simplicity. They placed brush-like, polymeric macromolecules on various solid and liquid surfaces to which the molecules' side-chains (the 'bristles') were strongly attracted. This attraction drove the bristles to spread out so as to maximize their contact with the surface, in turn causing the polymeric backbone of the molecule to stretch until it was eventually strained too far. Direct imaging of the size of the molecules using atomic force microscopy proved that they had been torn apart, just as if the rope had failed in a game of tug of war (Fig. 1).

When a chemical bond snaps, a chemical

reaction takes place. But exactly how can a purely mechanical effect have chemical consequences? That is the wider question investigated by a field known as chemomechanics². Future research in this area might well focus on whether mechanical stress shifts the energy level of the transition state of a chemical

reaction, through which reactants must pass before becoming products. This view has long been held for 'hard' materials, such as metals, ceramics and semiconductors (steel, for example, corrodes most easily when stressed and bent³). Mechanical stress could thus lower the activation threshold of certain reactions, making it easier to kick-start them, or even, depending on the magnitude of the stress applied, switch them off and on.

Organic systems such as that of Sheiko *et al.*¹ add a new twist: whereas the basic unit of hard materials is the atom, that of an organic material is the molecule. The added complexity of a molecule's internal architecture means that the kind of stress transmission described by the authors could not have been observed in an atomic system. In biology, the fact that a mechanical stimulus has chemical consequences — for example, in cellular processes that sense mechanical change⁴ or changes in the conformation and function of ion channels in lipid membranes⁵ — is being increasingly acknowledged. Thus, investigations of the effect of stress on complex molecules have considerable appeal.

In many organic materials and elastomers (rubbers or rubber-like plastic), the internal architecture of the molecule focuses large stresses on weaker chemical bonds, and stress-induced scission of chemical bonds in such materials is a costly problem⁶. We believe that the work of Sheiko and colleagues¹, by pointing the way towards understanding this ubiquitous and deleterious phenomenon, could provide a general model for designing molecular materials that have an architecture better able to cope with mechanical stress. Such research has myriad technological implications, because the ideas suggested here comprise a new paradigm for solving those problems.

What is in our opinion even more exciting is that there is a general proof-of-concept here — that slow or even forbidden chemical reactions can be activated by mechanical stress. Chemical reactivity clearly depends on the relative orientation of the reactants; so could mechanical deformation be used to place molecules in

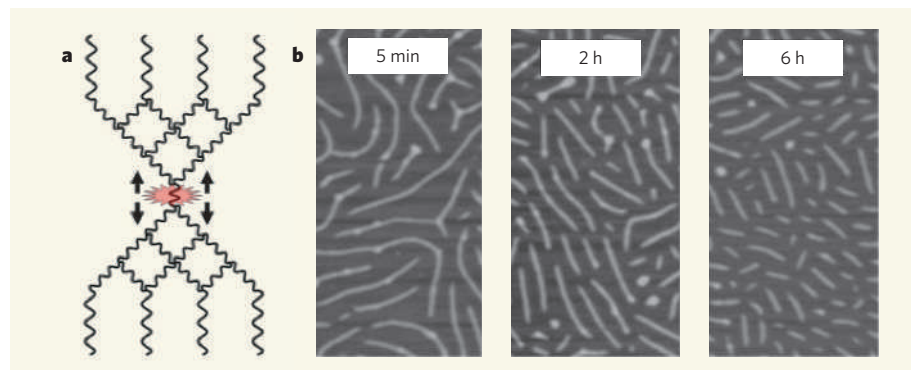


Figure 1 | Bond breaking. **a**, Channelling mechanical stress to specific, weaker chemical bonds can trigger chemical reactions that otherwise occur only with great effort. **b**, Sheiko *et al.*¹ implemented this idea with brush-like macromolecules on a solid surface. As the bristles of these brushes spread to maximize their contact with the surface, the resultant force is concentrated at the middle, causing chemical bonds to break: the molecules' length thus decreases with time.

a more favourable alignment for reaction? For example, friction and confinement offer a versatile way to align molecules⁷, and tribologists, who earn their bread from the study of these things, have known for a long time that friction promotes chemical reactions⁸. The tribological question^{7,8} is not directly addressed in Sheiko and colleagues' experiments, but will be an interesting motivation for further work.

And what about using the heat released to accomplish useful chemical change? Carbon-carbon bonds, by dint of their strength, release a lot of energy when they are broken — just as, in a bout of tug of war, energy is lost when the rope fails and competing teams fall to the ground. Although Sheiko *et al.* did not set out to capture energy from breaking carbon-carbon bonds, there is no reason that molecules could not be designed that use this energy for productive chemical means. In

this area, too, Sheiko *et al.*¹ have presented a fundamental proof-of-principle on which further efforts can be built and go beyond the more obvious, unwanted consequences of mechanical degradation.

Steve Granick and Sung Chul Bae are in the Departments of Materials Science and Engineering, and Chemistry and Physics, University of Illinois, Urbana, Illinois 61801, USA. e-mails: sgranick@uiuc.edu; sungbae@uiuc.edu

1. Sheiko, S. S. *et al.* *Nature* **440**, 191–194 (2006).
2. Beyer, M. K. & Clauser-Schaumann, H. *Chem. Rev.* **105**, 2921–2948 (2005).
3. Balluffi, R. W., Allen, S. M. & Carter, W. C. *Mechanics of Materials* (Wiley, New York, 2005).
4. Kung, C. *Nature* **436**, 647–653 (2005).
5. Wiggins, P. & Phillips, R. *Biophys. J.* **88**, 880–902 (2005).
6. Porter, R. S. & Casale, A. *Polym. Eng. Sci.* **25**, 129–156 (1985).
7. Bae, S. C., Lee, H., Lin, Z. & Granick, S. *Langmuir* **13**, 5685–5688 (2005).
8. Heinicke, G. *Tribochimistry* (Akademie, Berlin, 1984).

COMPARATIVE GENOMICS

Difference of expression

Rasmus Nielsen

Evolutionary studies tend to focus on alterations in proteins. But evolutionary change can often occur through modified gene expression, a process that is now under investigation with species-specific microarrays.

Gene expression, the major determining factor of protein abundance in the cell, is regulated by various mechanisms, such as protein binding to the DNA sequence and interference by small RNA molecules. On page 242 of this issue¹, Gilad *et al.* describe their study of gene expression in four primates. Their work is aimed at identifying similarities and differences in gene expression between humans and their nearest relatives.

As we elucidate the complex molecular machinery that controls gene expression, our ignorance of its role in evolution is becoming increasingly alarming. In most cases, we know little about the way in which gene expression is involved in how organisms adapt to new environments or otherwise evolve. It has long been hypothesized that adaptation over short evolutionary time may often proceed by modifications in the regulation and interaction of genes rather than in the protein gene-products themselves². Proteins tend to interact in complex networks, and so small changes in the abundance of one protein may have profound consequences. At the DNA level there may be many different mutations that affect gene-expression levels, but very few potentially beneficial mutations that directly affect protein function. Nonetheless, for convenience, most evolutionary studies have focused on protein evolution, leaving gene expression as one of the great unknowns in evolutionary biology.

Gilad *et al.*¹ make new strides in this field of

research. They compare gene-expression data in humans, chimpanzees, orangutans and rhesus monkeys to identify genes that have changed their level of expression in the human lineage. This research differs from earlier studies^{3,4} in using microarrays designed specifically for each species. Microarrays consist of a number of probes that bind messenger RNA from specific genes (mRNA is the linking molecule between a gene and the protein it encodes). By determining how much mRNA binds to each probe, the relative abundance of mRNA from each gene can be assessed. However, if the same microarray is used for all species, results may differ between species because of species-specific mutations that affect the binding affinity of the probes. Although this problem can be partially circumvented by removing genes that have such mutations, species-specific microarrays are the only known way to obtain a fair comparison among several divergent species without the loss of any genes.

Using this technology, Gilad and colleagues demonstrate that most genes are under natural selection to maintain a constant level of expression, but that a few genes show evidence of species-specific changes. The fact that selection in most cases is working to maintain expression levels near some optimum is not surprising — levels of expression of a gene that are too high or too low would presumably often be detrimental to an organism. Gilad

et al. also observe no systematic increase or decrease in the regulation of gene expression in either humans or chimpanzees, contrary to previous claims to this effect^{3–5}. But the authors do find that some groups of genes, particularly those encoding gene-transcription factors, tend to include greater numbers of upregulated genes in humans. Transcription factors are proteins that themselves play a role in regulating expression levels. So this observation is further support for the view that many evolutionary changes that are specific to humans may be related to gene expression.

Gilad *et al.*¹ also find that genes that are significantly up- or downregulated in humans, compared with other species, are often genes that have changed rapidly at the DNA-sequence level⁶. So there seems to be a correspondence between genes with altered expression and genes that have been targeted by positive darwinian selection in their protein-coding regions. This makes sense — we would expect changes in the function of a protein to be followed by changes in its distribution and abundance. Likewise, we may expect genes that have suffered a loss or reduction in functionality to subsequently experience an increased rate of evolution in both the sequence of the protein it encodes and its expression level, because selective constraints on it will have been relaxed.

What factors might be causing differences in gene expression between species? Such factors could include changes in the DNA close to the gene, for example changes in transcription-factor binding sites, or in distantly located elements such as gene enhancers, RNA genes or genes encoding transcription factors. Quantifying the relative importance of the evolution of these various elements will not be easy, but large-scale studies comparing many different organisms should reveal correlations between evolutionary changes at the DNA level and changes in expression level or pattern. The comparative analysis of expression data may thereby serve to detect functional correlations between DNA and expression levels in organisms in which it is difficult to carry out direct studies using standard genetic techniques. The result, I predict, will be a new perception of the mechanisms underlying evolutionary change — one in which the emphasis is on changes in regulatory elements, in RNA genes and in segments of DNA other than protein-coding genes.

Rasmus Nielsen is at the Institute of Biology and the Centre for Bioinformatics, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen Ø, Denmark. e-mail: rasmus@binf.ku.dk

1. Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P. & White, K. P. *Nature* **440**, 242–245 (2006).
2. King, M. C. & Wilson, A. C. *Science* **188**, 107–116 (1975).
3. Enard, W. *et al.* *Science* **296**, 340–343 (2002).
4. Khaitovich, P. *et al.* *Science* **309**, 1850–1854 (2005).
5. Khaitovich, P. *et al.* *Genetics* **170**, 929–939 (2005).
6. Bustamante, C. D. *et al.* *Nature* **437**, 1153–1157 (2005).

OBITUARY

Richard Dalitz (1925–2006)

Particle physicist and creator of the Dalitz plot.

Richard Henry Dalitz was a giant in the field of particle physics. A theorist who always endeavoured to work close to experiment, his contributions over 60 years shed vital light on the nature of the fundamental forces and the constituents of matter.

Born in the small wheat-belt town of Dimboola, northwest of Melbourne, Australia, Dalitz gained degrees in both mathematics and physics at the University of Melbourne. Awarded a travelling scholarship, he moved in 1946 with his wife Valda to England, to study for a PhD under Nicholas Kemmer at Trinity College, Cambridge. There he benefited from the teaching of Paul Dirac, whose lecture course on quantum mechanics he attended twice. After two years, he ran short of funds and, with a small child to support, moved to a one-year post as assistant to Nevill Mott at the University of Bristol. In retrospect, Dalitz considered this year at Bristol vital to his subsequent research: there he first learned from the group of Cecil Powell about elementary particle physics and the 'strange' particles (so named because they left unusual tracks in the emulsions used to detect them) produced in cosmic-ray collisions. He started to think particularly about the nature of one of these, the 'tau meson', or K^+ , as it is now known.

Dalitz was invited by Rudolf Peierls to the University of Birmingham in 1949, where his encounters with Freeman Dyson proved most useful for his mastery of the quantum-electrodynamical methods that Richard Feynman was then developing to describe electromagnetic interactions. After completing his thesis, which was on transitions between spin states in the oxygen nucleus, Dalitz was finally able to concentrate on particle physics and study of the tau.

The result was two seminal contributions that bear his name: the study of the decay of the neutral pion (one of the lightest mesons, a class of particles now identified as quark–antiquark pairings) to a photon and an electron–positron, or 'Dalitz' pair; and the development of the 'Dalitz plot'. This plot presents the kinematic variables of the three-body final state of a reaction in two dimensions. This allowed so-called resonances — transient states that flag their existence through their decay to final-state particles of definite total energy — to be readily visible.

Dalitz was interested in the difference between the tau and another strange particle, the theta, which seemed to be the tau's identical twin except for the fact that

it decayed into two pions, whereas the tau decayed into three. Using his plot, he was able to establish that the tau, like the theta, has zero spin, so that the two particles could not be identical if parity holds. (Parity is the idea that reactions proceed the same way when all spatial coordinates are reversed.) Indeed, this 'tau–theta' puzzle was the first indication that parity did not hold for interactions involving the weak nuclear force. The subsequent realization by T. D. Lee and C. N. Yang that this could indeed be the case won them the 1957 Nobel Prize in Physics.

After brief periods at Cornell University in Ithaca, New York, and back at Birmingham, in 1956 Dalitz accepted a professorship at the Enrico Fermi Institute in Chicago. Working with Riccardo Levi-Setti, he developed what became a lifetime interest, often pursued in collaboration with Avraham Gal, in the recently discovered hypernuclei, in which a strange particle takes the place of a proton or neutron.

In 1963, Dalitz was persuaded by Peierls to return to Britain and join him at the University of Oxford as a Royal Society research professor and, from 1964, a fellow of All Souls College. At Oxford, Dalitz continued his work on the resonant states that were being discovered in ever-larger numbers — often through the use of Dalitz plots. These states could be grouped according to their mass, spin and parity into families of eight and ten. Murray Gell-Mann dubbed this the 'eightfold way', and proposed that the pattern could be explained if the resonances were combinations of fundamental building-blocks of fractional electric charge — quarks, as he named them.

At the time, most physicists considered quarks as merely a mathematical tool. Boldly, Dalitz pursued the possibility that they might be real dynamical objects, and, in a remarkable paper presented in Tokyo in 1965, showed how different combinations of three tightly bound 'up' and 'down' quarks explained many properties of the proton and neutron. He further proposed that the quark model could describe not only these ground states, but also the multitude of higher-energy resonant states.

At Oxford, Dalitz established a flourishing research programme to study the quark model, and attracted many senior physicists and students to it. Their work, and work worldwide, has shown that all particles found then and since fit well with quark-model predictions. This encompasses the original



strange particles, whose unusual properties are now attributed to the presence of at least one 'strange' quark in their make-up that hampers their decay, as well as states that contain the heavier quarks 'charm' and 'bottom'.

The quark model was the foundation of quantum chromodynamics (QCD), the comprehensive theory of the strong nuclear force that depicts quarks as being held together by the exchange of gluons — analogues of the photons of the electromagnetic force — which are coupled to a new 'colour' charge. Dalitz first heard of colour in a talk by Gell-Mann and recognized that it resolved some profound problems associated with the quark model. The success of this model once colour was included set the scene for the subsequent development of QCD. Dalitz's interest in the quark model continued all his working life. In 1992, with Gary Goldstein, he developed a method still in use today, to determine the detailed properties of the 'top' quark, the last of the six types of quark to be found.

A fellow of the Royal Society and recipient of many honours, among them the Hughes Medal and Royal Medal of the Royal Society, 'Dick' Dalitz never lost his appetite for physics. He remained an active member of the theoretical-physics department at Oxford after his retirement in 1990, always keen to encourage students to share in his passion for physics. He died on 13 January 2006, and is survived by his wife, son and three daughters.

Graham Ross

Graham Ross is in the Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.
e-mail: g.ross@physics.ox.ac.uk

G.-C. WICK/AIP

BRIEF COMMUNICATIONS

Nanomotor rotates microscale objects

A molecular motor in a liquid-crystal film uses light to turn items thousands of times larger than itself.

Nanomachines of the future will require molecular-scale motors^{1–6} that can perform work and collectively induce controlled motion of much larger objects. We have designed a synthetic, light-driven molecular motor that is embedded in a liquid-crystal film and can rotate objects placed on the film that exceed the size of the motor molecule by a factor of 10,000. The changes in shape of the motor during the rotary steps cause a remarkable rotational reorganization of the liquid-crystal film and its surface relief, which ultimately causes the rotation of submillimetre-sized particles on the film.

We used a specially designed motor (molecule **1** in Fig. 1a) featuring a right-handed helical structure and a single stereogenic centre in the (upper) rotor part that dictates the direction of rotation, a central carbon–carbon double bond that functions as an axle, and a (lower) stator part that resembles the liquid-crystal host⁷. Upon irradiation of motor molecule **1** with ultraviolet light of wavelength 365 nm, a photochemical isomerization around the central double bond occurs that results in inversion of the helicity (from right-handed to left-handed). A subsequent thermal step, again with helix inversion (left- to right-handed), occurs readily at 20 °C (with a reaction-time half-life of 9.9 min in toluene). Two photochemical steps, each followed by a thermal step, add up to a full 360° rotary cycle.

This motor is very effective at inducing helical organization in a liquid-crystal film. With its surface exposed to the air, a unidirectionally aligned cholesteric liquid-crystal film containing 1% by weight of molecule **1** as a dopant shows a polygonal fingerprint texture that is typical of cholesteric liquid crystals that have their helix axes parallel to the surface⁸ (Fig. 1b). When this sample is irradiated with light of wavelength 365 nm under an optical microscope, the polygonal texture reorganizes in a rotational (clockwise) fashion (for movie, see supplementary information). The rate of rotation gradually decreases until the process halts after about 10 min.

Removing the light source causes the rotation to resume, this time in the opposite direction. The textures always rotate clockwise during irradiation and anticlockwise during the thermal isomerization step. Exchanging molecule **1** for its enantiomer induces rotation in the opposite direction, confirming that the direction of rotation of the liquid-crystal tex-

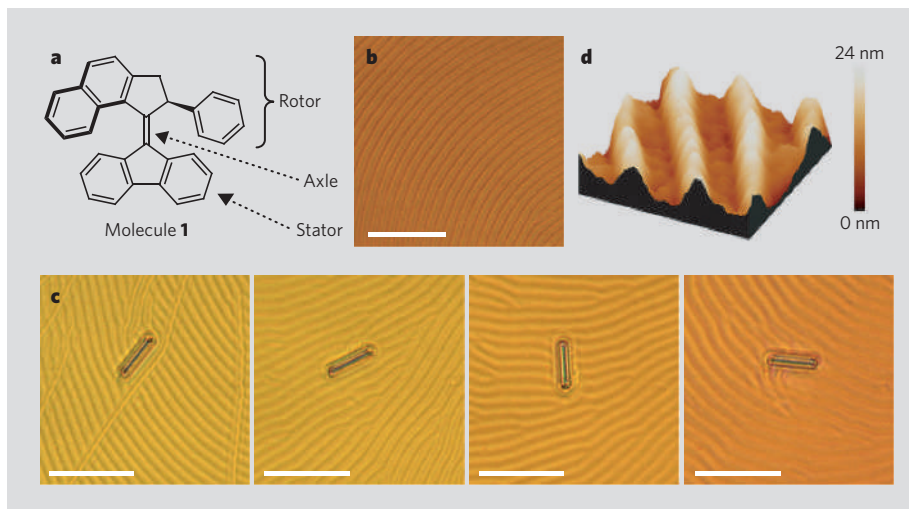


Figure 1 | Features of a light-driven molecular motor **a**, Structure of the motor (molecule **1**). Bonds in bold point out of the page. **b**, Polygonal texture of a liquid-crystal film doped with molecule **1** (1% by weight). **c**, Glass rod rotating on the liquid crystal during irradiation with ultraviolet light. Frames 1–4 (from left) were taken at 15-s intervals and show clockwise rotations of 28° (frame 2), 141° (frame 3) and 226° (frame 4) of the rod relative to the position in frame 1 (for movies, see supplementary information). Scale bars, 50 μm. **d**, Surface structure of the liquid-crystal film (atomic force microscopy image; 15 μm²).

ture is determined by the change in helicity of the motor.

The rotation of the texture induced by the motor can be harnessed to move submillimetre-sized particles placed on top of the film. Figure 1c shows the first stages (over 45 s) of a typical rotary motion of a microscopic glass rod (for movie, see supplementary information). The glass rod (5 × 28 μm) rotates in the same direction as the cholesteric texture during the photochemical and thermal steps of the motor at an average speed of 0.67 and 0.22 r.p.m., respectively.

Using non-contact atomic force microscopy, we found that the liquid-crystal film doped with molecule **1** has a surface relief that is 20 nm in height (Fig. 1d). Optical profilometry indicates that the orientation of this surface relief alters in response to photochemically or thermally induced topology changes in the embedded molecular motor (see methods in supplementary information). This reorganization generates a torque on the microscopic object that results in rotary motion.

We have described a collective change in helicity in a nanosized motor that can be used to rotate microscopic-scale objects by harvesting light energy, and have demonstrated that a rotary molecular motor can perform work.

Rienk Eelkema*, Michael M. Pollard*,

Javier Vicario*, Nathalie Katsonis*, Blanca Serrano Ramon†, Cees W. M. Bastiaansen†, Dirk J. Broer‡, Ben L. Feringa*

*Department of Organic and Molecular Inorganic Chemistry, Stratingh Institute, University of Groningen, 9747 AG Groningen, The Netherlands e-mail: b.l.feringa@rug.nl

†Department of Polymer Technology, Faculty of Chemistry and Chemical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

‡Philips Research Laboratories, Department of Biomolecular Engineering, High Tech Campus 4, 5656 AE Eindhoven, The Netherlands

- Schliwa, M. (ed.) *Molecular Motors* (Wiley, Weinheim, 2003).
- Balzani, V., Venturi, M., Credi, A. *Molecular Devices and Machines: A Journey into the Nanoworld* (Wiley, Weinheim, 2004).
- Kelly, T. R., De Silva, H. & Silva, R. A. *Nature* **401**, 150–152 (1999).
- Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. *Nature* **401**, 152–155 (1999).
- Leigh, D. A., Wong, J. K. Y., Dehez, F. & Zerbetto, F. *Nature* **424**, 174–179 (2003).
- Liu, Y. et al. *J. Am. Chem. Soc.* **127**, 9745–9759 (2005).
- Koumura, N., Geertsema, E. M., Meetsma, A. & Feringa, B. L. *J. Am. Chem. Soc.* **122**, 12005–12006 (2000).
- Dierking, I. *Textures of Liquid Crystals* (Wiley, Weinheim, 2003).

Supplementary Information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Received 8 September 2005; accepted 1 February 2006.
doi:10.1038/440163a

GAMMA-RAY BURSTS

Huge explosion in the early Universe

Long gamma-ray bursts (GRBs) are bright flashes of high-energy photons that can last for tens of minutes; they are generally associated with galaxies that have a high rate of star formation and probably arise from the collapsing cores of massive stars, which produce highly relativistic jets (collapsar model¹). Here we describe γ - and X-ray observations of the most distant GRB ever observed (GRB 050904): its redshift^{2,3} (z) of 6.29 means that this explosion happened 12.8 billion years ago, corresponding to a time when the Universe was just 890 million years old, close to the reionization era⁴. This means that not only did stars form in this short period of time after the Big Bang, but also that enough time had elapsed for them to evolve and collapse into black holes.

GRB 050904 triggered the Burst Alert Telescope (BAT) on board the Swift⁵ satellite on 4 September 2005 at 1:51:44 GMT. The spacecraft quickly slewed to allow observations by

the X-ray Telescope (XRT)^{6,7}, which measured the burst for ten days after its onset. Figure 1 (top panel) shows the history of the burst. We shall present and discuss the GRB phenomenology from the point of view of the rest frame of its source.

The BAT light curve shows three main peaks: two of about 2 s at $T + 3.8$ s and $T + 7.7$ s, and a long-lasting one at about $T + 13.7$ s, where T is the time of the burst onset. It also shows a weak peak at about $T + 64$ s. The early XRT light curve shows a steep power-law decay with an index of -2.07 ± 0.03 ; two flares are superimposed at $T + 64$ s (coincident with the last peak of the BAT light curve) and $T + 170$ s. Although interrupted by the constraints of low-Earth-orbit observation, the X-ray light curve reveals highly irregular intensity variations, probably due to the presence of flares for up to $T + 1.5$ hours. At later times, flaring activity is not detected, leaving only a residual emission that is 10^6 times lower than the initial intensity.

The flares in the XRT light curve can be interpreted as late internal shocks related to central engine activity. In this scenario, they would have the same origin as the first γ -ray emission^{8–10}, which would require the central engine to remain active for at least 5,000 seconds, consistent with the collapsar model¹.

Spectral analysis was performed by selecting time intervals corresponding to characteristic phases of the light curve evolution. All spectra were well modelled by a single power law, with both galactic and intrinsic absorption components in the case of the XRT spectra. Figure 1 (bottom) shows the evolution with time of the photon index Γ . The BAT spectra have $\Gamma = -1.2$. If we exclude the spectrum of the first XRT flare at $T + 64$ s, the XRT photon indices show a clear, decreasing trend from about -1.2 to about -1.8 in the first $T + 200$ s. No further spectral evolution is present in later XRT data.

The overall phenomenology of GRB 050904 is not peculiar with respect to other GRBs at lower redshift. This suggests that the mechanisms of GRB explosions in the early Universe and today are similar.

Based on the likely existence of population I/II stars in galaxies that were already metal-enriched at these high redshifts¹¹, we expect about 10% of all bursts detected by Swift to be located at $z \geq 5$. A higher percentage would require an additional contribution to the high-redshift GRB population by metal-free population III stars, which are viable GRB progenitors for long-duration GRBs¹¹. A more systematic search for GRB optical counterparts will increase the sample of these high-redshift GRBs, allowing us to probe the existence of metal-free massive stars of population III.

G. Cusumano¹, V. Mangano¹, G. Chincarini^{2,3}, A. Panaitescu⁴, D. N. Burrows⁵, V. La Parola¹,

T. Sakamoto^{6,7}, S. Campana², T. Mineo¹, G. Tagliaferri², L. Angelini⁶, S. D. Barthelémy⁶, A. P. Beardmore⁸, P. T. Boyd⁶, L. R. Cominsky⁹, C. Gronwall⁵, E. E. Fenimore⁴, N. Gehrels⁶, P. Giommi¹⁰, M. Goad⁸, K. Hurley¹¹, J. A. Kennea⁵, K. O. Mason¹², F. Marshall⁶, P. Mészáros^{5,13}, J. A. Nousek⁵, J. P. Osborne⁸, D. M. Palmer⁴, P. W. A. Roming⁵, A. Wells⁸, N. E. White⁶, B. Zhang¹⁴

¹INAF-Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo, 90146 Palermo, Italy

e-mail: cusumano@ifc.inaf.it

²INAF Osservatorio Astronomico di Brera, 23807 Merate, Italy; ³Università degli studi di Milano-Bicocca, Dipartimento di Fisica, 20126 Milan, Italy; ⁴Los Alamos National Laboratory, PO Box 1663, Los Alamos, New Mexico 87545, USA; ⁵Department of Astronomy & Astrophysics and

¹³Department of Physics, Pennsylvania State University, Philadelphia, Pennsylvania 16802, USA; ⁶NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA; ⁷National Research Council, TJ2114, Washington DC 20418, USA; ⁸Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK; ⁹Department of Physics and Astronomy, Sonoma State University, Rohnert Park, California 94928-3609, USA; ¹⁰ASI Science Data Center, 00044 Frascati, Italy; ¹¹Space Sciences Laboratory, University of California, Berkeley, California 94720-7450, USA; ¹²Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK; ¹⁴Department of Physics, University of Nevada, Box 454002, Las Vegas, Nevada 89154-4002, USA

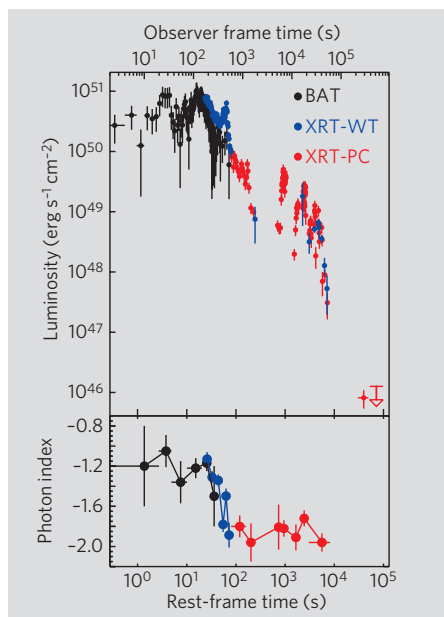


Figure 1 | Light curve and spectral evolution of GRB 050904 as observed by the Burst Alert Telescope (BAT) and the X-ray Telescope (XRT). WT, windowed timing mode data; PC, photon counting data. Top, evolution of the gamma-ray burst (GRB) K -corrected 0.2–10 keV luminosity (the K -correction accounts for the redshift dependence of the luminosity in a given band of wavelengths). Error bars show 90% confidence. Times are referred to the BAT trigger. Rest-frame time is obtained by applying correction factor $(1+z)^{-1}$ to the observer frame time. Gaps in XRT-PC data correspond to the part of the orbit when the satellite was not observing this GRB. Bottom, change in photon index (Γ , defined by the power law $F(E) = E^{\Gamma+1}$, where $F(E)$ is the observed flux of energy E) of GRB 050904 during the observation. Spectra were modelled using a power law with two absorbing components (galactic and intrinsic).

- MacFadyen, A. I., Woosley, S. E. & Heger, A. *Astrophys. J.* **550**, 410–425 (2001).
- Tagliaferri, G. et al. *Astron. Astrophys.* **443**, L1–L5 (2005).
- Kawai, N. et al. *GCN Circ.* 3937 (2005).
- Becker, R. H. et al. *Astron. J.* **122**, 2850–2857 (2001).
- Gehrels, N. et al. *Astrophys. J.* **611**, 1005–1020 (2004).
- Sakamoto, T. et al. *GCN Circ.* 3938 (2005).
- Mineo, T. et al. *GCN Circ.* 3920 (2005).
- Burrows, D. N. et al. *Science* **309**, 1833–1835 (2005).
- Zhang, B. et al. *Astrophys. J.* (in the press); preprint at <<http://arXiv.org/astro-ph/0508321>> (2005).
- Nousek, J. A. et al. *Astrophys. J.* (in the press); preprint at <<http://arXiv.org/astro-ph/0508332>> (2005).
- Bromm, V. & Loeb, A. *Astrophys. J.* (in the press); preprint at <<http://arXiv.org/astro-ph/0509303>> (2005).

Competing financial interests: declared none.

Received 11 November 2005; accepted 23 November 2005.
doi:10.1038/440164a

CORRIGENDUM

Developmental technology: Dogs cloned from adult somatic cells

Byeong Chun Lee, Min Kyu Kim, Goo Jang, Hyun Ju Oh, Fibrianto Yuda, Hye Jin Kim, M. Hossein Shamim, Jung Ju Kim, Sung Keun Kang, Gerald Schatten, Woo Suk Hwang
Nature **436**, 641 (2005)

Supplementary Table 1 of this communication has been replaced (9 March 2006; corrections shown in red) as the original peak values reported for two of the canine microsatellite markers (PEZ02, REN105L03) were in error; also, those for PEZ08 have been removed (further details are available from B.C.L. at bclee@snu.ac.kr). In the Table legend, the URL giving details of the markers has been updated. The patent application mentioned in the legend should originally have been declared as a competing financial interest.
doi:10.1038/440164b

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see *Nature* contents.

MOLECULAR GENETICS

DNA analysis of a putative dog clone

Arising from: B. C. Lee *et al.* *Nature* **436**, 641 (2005)

In August 2005, Lee *et al.* reported the first cloning of a domestic dog from adult somatic cells¹. This putative dog clone was the result of somatic-cell nuclear transfer from a fibroblast cell of a three-year-old male Afghan hound into a donor oocyte provided by a dog of mixed breed. In light of recent concerns regarding the creation of cloned human cell lines from the same institution^{2,3}, we have undertaken an independent test to determine the validity of the claims made by Lee *et al.*¹.

Duplicate sets of blood samples were provided from the original fibroblast donor dog (Tai, an Afghan hound), the surrogate mother (a Labrador retriever) and Snuppy, the putative clone. Samples were drawn in heparinized tubes and delivered to us on ice overnight.

Collection and mailing of samples was supervised by In Kwon Chung, a member of the investigative committee at Seoul National University in South Korea. Samples were not provided from the oocyte donor, which was unavailable for sampling. Samples were coded by a third party and laboratory personnel were blind to sample identifiers.

In addition to these six samples, the test panel included previously purified DNA samples from 11 Afghan hounds collected in the United States and registered with the American Kennel Club (AKC); eight of these shared no common parents or grandparents, and the other three were half-siblings that shared a common sire. Pedigree analysis revealed that one of the American-collected dogs was a first

cousin of Tai, the fibroblast donor, and that five of the others had distant maternal and paternal relatives in common with him. Other samples on the test panel included a pure-bred female Labrador retriever, purportedly unrelated to the surrogate mother, and Tasha, the boxer dog used for the reference canine sequence⁴.

We tested both nuclear and mitochondrial markers. Nuclear markers included 16 microsatellite markers routinely used for canine paternity testing by the AKC^{5,6}. (A seventeenth marker was discarded because it failed to amplify from 25% of the DNA samples.) For all nuclear markers tested, Snuppy and Tai, the clone and donor, had identical genotypes (Table 1).

Table 1 | Microsatellite genotypes of 16 dogs, including the putative clone, the nuclear-DNA donor and the surrogate mother

	FH2010	FH2054	FH2079	Pez01	Pez05	Pez06	Pez08	Pez10	Pez11	Pez12	Pez13	Pez15	Pez16	Pez17	Pez20	Pez21
Boxer	242	178	292	134	120	199	244	306	161	293	235	224	319	226	191	109
	242	178	292	138	124	202	244	306	161	315	243	240	323	230	194	109
Labrador	246	166	288	130	128	187	256	310	153	285	239	228	311	222	194	109
	254	174	288	138	128	187	264	310	165	285	243	236	327	226	194	120
Afghan-1	246	170	288	138	120	198	252	283	145	300	243	228	315	226	194	109
	246	186	288	138	120	203	252	310	145	315	250	232	335	226	198	112
Afghan-2	246	174	288	130	120	194	252	283	145	289	nd	232	315	226	198	109
	254	186	288	138	120	198	252	298	145	320	nd	232	335	226	198	116
Afghan-3	246	174	292	113	120	199	252	283	145	289	243	232	315	226	198	112
	250	186	308	138	120	199	252	298	153	326	243	232	327	226	198	120
Afghan-4	246	186	288	113	124	199	252	302	150	293	250	228	335	226	194	112
	246	186	288	113	132	203	252	302	157	300	250	228	335	226	198	116
Afghan-5	246	174	288	130	120	194	252	298	145	315	235	228	315	226	198	109
	250	186	308	138	120	199	264	298	157	326	243	232	335	226	198	116
Afghan-6	246	174	292	138	120	195	252	283	145	275	235	232	315	226	194	112
	246	194	292	138	124	203	252	310	150	293	243	232	335	230	198	116
Afghan-7	246	174	288	130	132	203	252	298	150	289	219	228	315	226	194	112
	258	190	288	138	132	207	252	322	153	289	250	232	319	230	194	116
Afghan-8	246	174	288	113	120	194	252	294	157	289	219	232	311	226	198	116
	246	186	292	138	124	203	256	310	166	315	243	240	315	226	198	116
Afghan-9	246	186	282	138	124	199	252	294	166	315	219	228	315	230	198	109
	254	194	308	138	124	203	264	294	166	315	219	228	315	234	198	116
Afghan-10	246	186	288	113	132	195	241	283	157	312	219	232	323	230	198	112
	254	186	292	138	132	199	252	294	166	326	252	232	335	230	198	116
Afghan-11	246	174	nd	138	120	195	252	326	153	nd	243	228	315	226	194	109
	246	194	nd	138	132	198	252	nd	157	nd	250	228	330	226	198	116
Donor	246	186	288	130	124	199	252	294	153	300	235	232	307	226	194	112
	246	194	288	138	132	199	252	298	153	326	235	232	315	230	194	116
Snuppy	246	186	288	130	124	199	252	294	153	300	235	232	307	226	194	112
	246	194	288	138	132	199	252	298	153	326	235	232	315	230	194	116
Surrogate	250	170	292	138	120	198	256	298	145	285	215	228	311	230	194	116
	254	170	296	138	124	203	256	298	153	297	243	236	323	234	198	120

Alleles are named for the total length of the segment amplified; nd, allele not determined.

Methods. DNA was extracted and purified from all blood samples using the QIAamp DNA blood mini kit (Qiagen). The conditions and protocol used for the polymerase chain reaction are available at http://research.nhgri.nih.gov/dog_genome. Amplicons were genotyped using an ABI 3730xl DNA analyser. For all 19 samples, a 614-base segment of the mitochondrial DNA control region was sequenced using ABI BigDye terminator chemistry and standard protocols (see http://research.nhgri.nih.gov/dog_genome for details and raw data). Primers for both amplification and sequencing were designed using publicly available information (www.genome.ucsc.edu) and were: forward, 5'-TGAATCACCCCTACTGTGCTATGT-3' and reverse, 5'-ACCTTGATTATATCGCTGAGTTGA-3'. All markers and sequences were run in duplicate in independent assays, yielding more than 99% of potential genotypes (see also Table 3).

Table 2 | Probability of exact allelic matching at 16 microsatellite loci in a population of Afghan hounds

Scenario	Probability
Hardy–Weinberg*	7×10^{-14}
Full siblings†	9×10^{-6}
Inbred $F = 0.25$ ‡	3×10^{-8}
Pedigree§	4×10^{-4}

*A random sampling of the population, assuming Hardy–Weinberg conditions.
†Assuming the clone is a full sibling of the donor.
‡Assuming everyone in the population is related with an inbreeding coefficient of 0.25.
§An assumed pedigree in which the donor is the product of a mother–son mating, and the putative clone is the product of crossing the donor back to the mother.

The probability that the putative clone should have precisely the same genotype as the donor was computed for different assumptions regarding the relatedness of the sample and the donor⁷ (Table 2). In all cases, the allele frequencies were computed from a sample of 11 AKC-registered Afghan hounds plus the donor. According to genetic maps of the canine chromosomes, the 14 mapped markers were unlinked to one another, so each microsatellite was treated as an independent locus^{8,9}.

The match probabilities ranged from 7×10^{-14} for unrelated dogs to 4×10^{-4} for

those with a specific inbred pedigree. A higher degree of inbreeding would increase the match probability further, but the donor does not seem to be extremely inbred; both the donor and Snuppy are heterozygous at 8 of the 16 markers, which is not significantly different from the number of heterozygous markers expected, given the observed allele frequencies and no inbreeding.

Mitochondrial D-loop sequencing revealed 26 variable bases within the 614 analysed (Table 3). Snuppy and the donor dog differed at 12 of the 26 sites. Nine of the Afghan hound sequences disagreed with each other at only one base (position 548) and differed from the donor by only three to four bases. Also, the two Labrador retrievers had identical mitochondrial sequences that differed from the donor by only three bases. The sequence from Snuppy differed from that of any other dog at 9–14 sites.

These data are consistent with Snuppy being a genetic clone of the donor dog Tai. Our analysis rules out most feasible alternatives to a true clone, such as the production of a delayed twin, which would have produced dogs with the same mitochondrial D-loop sequence, or an animal resulting from extreme inbreeding, which would have yielded dogs

that were homozygous at more than the observed eight loci.

Conclusions drawn from these results are subject to caveats. First, we did not witness the drawing of the blood samples, which was done under the supervision of a third party. However, no obvious hypothetical manipulation of the samples would have generated the results described here — perfect matching of the nuclear markers, and distinct differences between the mitochondrial sequences for the donor and Snuppy. Second, we were not provided with samples from the oocyte donor, although tissue samples from this dog have been tested by investigators at Seoul National University¹⁰. Without this sample, we are unable to confirm the original experimental details¹, or to say with certainty that the mitochondrial variants observed were those that were expected.

Finally, our statistical analysis is based on a limited number of Afghan hounds. A larger number of unrelated individuals might have provided a more precise estimate of population allele frequencies. However, given that the dogs tested are representative of this relatively restricted breed and that many share a partial heritage with the donor, the statistical conclusions are conservative.

Heidi G. Parker*, Leonid Kruglyak†, Elaine A. Ostrander*

*Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA
e-mail: eostrand@mail.nih.gov
†Lewis-Sigler Institute for Integrative Genomics, and the Department of Ecology and Evolutionary Biology, Carl Icahn Laboratory, Princeton University, Princeton, New Jersey 08544, USA

Table 3 | Mitochondrial genotypes at 26 polymorphic loci within the canine mitochondrial D-loop

Loci*	49	76	110	118	135	143	148	150	153	155	159	162	166	175	304	323	334	338	371	435	478	482	526	548	561	562
Boxer	C	A	T	C	T	C	T	G	C	C	T	A	A	G	C	T	A	T	G	C	T	C	A	T	C	T
Labrador	C	A	T	C	T	C	T	G	C	C	T	A	A	G	C	T	A	T	G	C	C	C	A	T	C	T
Afghan-1	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	C	C	T
Afghan-2	C	A	T	C	T	T	C	A	T	T	C	A	A	G	C	C	A	C	A	T	C	T	G	T	C	C
Afghan-3	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	T	C	T
Afghan-4	T	A	T	T	C	T	T	A	C	T	T	G	G	A	C	C	A	C	G	T	T	C	G	T	C	T
Afghan-5	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	T	C	T
Afghan-6	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	C	C	T
Afghan-7	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	T	C	T
Afghan-8	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	C	C	T
Afghan-9	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	C	C	T
Afghan-10	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	C	C	T
Afghan-11	C	A	T	C	T	T	T	A	C	C	T	A	A	G	T	T	A	T	G	C	C	C	A	C	C	T
Donor	C	A	T	C	T	T	T	G	C	C	T	A	A	A	C	T	A	T	G	C	C	C	A	C	C	T
Snuppy	T	G	C	C	T	T	T	A	C	T	T	A	A	A	C	C	G	T	G	T	T	C	G	T	T	T
Surrogate	C	A	T	C	T	C	T	G	C	C	T	A	A	G	C	T	A	T	G	C	C	C	A	T	C	T

*The locations of the variable bases are given relative to the 5' end of the forward primer used to amplify and sequence the region. The entire length of the segment is 614 bases.

MOLECULAR GENETICS

Verification that Snuppy is a clone

Arising from: B. C. Lee *et al. Nature* **436**, 641 (2005)

Somatic-cell nuclear-transfer technology has been used to clone a variety of animal species^{1–3}, but the overall efficiency of the cloning process and the viability of embryos has remained low⁴. Until Lee *et al.* described the cloning of two Afghan hounds by nuclear transfer from adult skin fibroblasts into oocytes that had matured

*in vivo*⁵, dog cloning had been unsuccessful because of the difficulty of collecting canine oocytes matured *in vivo* at metaphase II (ref. 6). Here we provide independent evidence from the Seoul National University Investigation Committee that Snuppy, the one of the pair to survive, is a genuine clone.

To investigate whether the cloned dog was genetically identical to the donor Afghan, we obtained blood samples from Snuppy, from the male Afghan hound that provided the somatic cell, and from the surrogate mother. In addition, autopsy samples from the since-deceased mixed-breed dog that originally provided the egg used to create Snuppy were obtained from the research team who generated the cloned dog. DNA was extracted from the blood and autopsy samples and used for microsatellite analysis of genomic DNA and nucleotide sequences of mitochondrial DNA.

Table 1 | Analysis of canine-specific microsatellite loci

Canine markers*	Snuppy (blood leukocytes)		Donor Afghan (blood leukocytes)		Surrogate (blood leukocytes)		Egg donor* (autopsy lung tissue)	
	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2
AHT111	72	78	72	78	78	80	76	80
AHT137	130	148	130	148	130	148	130	142
C01.424	186	188	186	188	186	188	182	188
C07.620	106	114	106	114	112	114	114	124
Wilms-TF	291		291		291		297	299
FH2274	278	286	278	286	278		278	282
FH2289	284	328	284	328	296		288	292
REN210K18	216		216		212	217	202	217

*Eight of 27 canine-specific markers used for microsatellite assay are shown.

Table 2 | Sequence alignments within D-loop region, cytochrome *b*, and 16S rRNA of mtDNA

	Nucleotide positions											
	D-loop region (409 bp)									Cytochrome <i>b</i> (380 bp)	16S rRNA (440 bp)	
	170	175	195	343	354	455	498	546	568	426	1117	1145
Snuppy	A	T	A	C	G	T	T	G	T	G	C	G
Donor Afghan	G	C	A	T	A	C	C	A	C	A	T	A
Surrogate	G	C	G	T	A	C	C	A	T	A	C	A
Egg donor	A	T	A	C	G	T	T	G	T	G	C	G

*The nucleotide positions were counted from the start, the initiation codon, and the transcriptional start of the D-loop, cytochrome *b*, and 16S rRNA sequences, respectively (GenBank accession number AY729880). bp, Base pairs.

The microsatellite analyses were performed with genomic DNA from four dogs (Snuppy, the donor Afghan, the surrogate mother and the egg donor) using 27 canine-specific markers that allow extremely inbred animals to be distinguished. As shown in Table 1, which shows only 8 of 27 loci used for microsatellite analysis, Snuppy and the donor Afghan have identical microsatellite patterns for all loci, whereas the surrogate mother evidently has a different genetic origin.

For 18 of the microsatellite loci analysed (including those shown in Table 1), the egg donor and Snuppy showed distinct microsatellite patterns. The microsatellite loci used for the analyses show enough variation among Afghan hound individuals to allow them to be distinguished (results not shown). Thus, the Committee's microsatellite analysis of genomic

DNA demonstrates that the cloned dog Snuppy is genetically identical to its fibroblast-donor Afghan hound.

If Snuppy is a cloned dog, his mitochondrial DNA (mtDNA) should be identical to that of the egg donor, but not to that of the fibroblast-donor Afghan. To test this, we determined partial sequences of four regions of the dog mtDNA. Based on the complete nucleotide sequence (GenBank accession number U96639), four primer sets were synthesized and used for polymerase chain reaction (PCR)⁷: the cytochrome *b* gene (L14,252–L14,631), the 16S ribosomal RNA gene (L2,033–L2,472), and the two overlapping hypervariable D-loop regions (L15,622–L16,030 and L15,485–L16,090). PCR products were sequenced and a BLAST search confirmed their identity as mtDNA.

The sequence alignments of 409 base pairs

of the hypervariable D-loop region revealed eight mismatches between the mtDNA of Snuppy and the donor Afghan (Table 2). A non-match was observed even in more conserved regions (one mismatch in the cytochrome *b* gene and two mismatches in the 16S rRNA gene). These results indicate that Snuppy could not have been produced by splitting the early blastomere from which the donor Afghan originated. The sequence alignments also revealed a perfect match between Snuppy and the egg donor, and a non-match between Snuppy and the surrogate mother (Table 2). This is consistent with Snuppy being a nuclear clone of the donor Afghan.

On the basis of the results of both a microsatellite analysis of genomic DNA and a sequence comparison of mtDNA, it is highly unlikely that Snuppy came either from extreme inbreeding or from blastomere separation (also known as twinning). It is virtually certain that Snuppy was generated from somatic-cell nuclear transfer, as claimed by Lee *et al.*⁵.

Seoul National University Investigation Committee*, Jung Bin Lee†, Chankyu Park‡

*Office of Research Affairs, Seoul National University, Gwanak-gu, Seoul 151-742, South Korea
e-mail: mhchung@snu.ac.kr

†Department of Forensic Medicine, Seoul National University College of Medicine, 28 Yeongon-dong, Chongno-gu, Seoul 110-799, South Korea

‡Department of Life Sciences, Korea Advanced Institute of Science and Technology, Yusong-gu, Taejeon 305-701, South Korea

1. Wilmot, I. *et al. Nature* **385**, 810–813 (1997).
2. Wakayama, T. *et al. Nature* **394**, 369–374 (1998).
3. Zhou, Q. *et al. Science* **302**, 1179 (2003).
4. Sutovsky, P. & Prather, R. S. *Trends Biotech.* **22**, 205–208 (2005).
5. Lee, B. C. *et al. Nature* **436**, 641 (2005).
6. Westhusin, M. E. *et al. J. Reprod. Fert. Suppl.* **57**, 287–293 (2001).
7. Kim, K. S. *et al. Genes Genet. Syst.* **76**, 243–250 (2001).

doi:10.1038/nature04686

REVIEWS

Temperature sensitivity of soil carbon decomposition and feedbacks to climate change

Eric A. Davidson¹ & Ivan A. Janssens²

Significantly more carbon is stored in the world's soils—including peatlands, wetlands and permafrost—than is present in the atmosphere. Disagreement exists, however, regarding the effects of climate change on global soil carbon stocks. If carbon stored belowground is transferred to the atmosphere by a warming-induced acceleration of its decomposition, a positive feedback to climate change would occur. Conversely, if increases of plant-derived carbon inputs to soils exceed increases in decomposition, the feedback would be negative. Despite much research, a consensus has not yet emerged on the temperature sensitivity of soil carbon decomposition. Unravelling the feedback effect is particularly difficult, because the diverse soil organic compounds exhibit a wide range of kinetic properties, which determine the intrinsic temperature sensitivity of their decomposition. Moreover, several environmental constraints obscure the intrinsic temperature sensitivity of substrate decomposition, causing lower observed 'apparent' temperature sensitivity, and these constraints may, themselves, be sensitive to climate.

The temperature sensitivity of decomposition of the enormous global stocks of soil organic matter (SOM)^{1–4} has recently received considerable interest, including several high-profile publications supporting opposing views^{5–14}. Interest in this topic is high because of its importance in the global carbon (C) cycle and potential feedbacks to climate change¹⁵. This recent controversy has focused primarily on organic matter in upland mineral soils. These soils have reasonably good drainage and aeration, allowing roots and soil fauna to penetrate into mineral soil layers, thus mixing SOM with mineral particles. Conditions in upland mineral soils are also generally favourable for decomposition, resulting in relatively low carbon densities. In contrast, in wetlands and peatlands where anaerobic conditions frequently persist, decomposition proceeds much more slowly, and deep layers of organic matter accumulate on top of mineral layers. In soils with permanently frozen layers (permafrost), drainage is also often poor, and organic matter may become buried in deep soil layers through cryoturbation¹⁶. Thus, wetlands, peatlands and permafrost soils generally contain higher carbon densities than upland mineral soils, and together they make up enormous stocks of carbon globally (Table 1). Moreover, permafrost soils and a large fraction of peatland soils occur at high latitudes, where warming is expected to be greatest, and, indeed, has already begun^{17,18}. In this review, we emphasize that decomposition of all types of belowground organic matter should be described by a common set of principles of kinetic theory and environmental constraints. Our objective is to clarify the issues regarding temperature sensitivity of decomposition within a framework that helps to focus the ensuing debate and research.

Factors controlling decomposition of organic matter

The stocks of organic matter in soils result from the balance between inputs and outputs of carbon within the belowground environment (Fig. 1). Inputs are primarily from leaf and root detritus. Outputs are

dominated by the efflux of carbon dioxide (CO₂) from the soil surface, although methane (CH₄) efflux and hydrologic leaching of dissolved and particulate carbon compounds can also be important. The production of CO₂ in soils is almost entirely from root respiration and microbial decomposition of organic matter. Like all chemical and biochemical reactions, these processes are temperature-dependent. Root respiration and microbial decomposition are also subject to water limitation. Hence, most empirical models relate the efflux of CO₂ from soils (often lumping microbial and root respiration together as 'soil respiration') to temperature and often also to some scalar of soil water content or precipitation^{19–24}. This much is not controversial.

The kinetics of enzymatic reactions in well-mixed media are also not controversial. Activation energies are related to the ambient temperature and to the molecular structure of the organic-C reactant. The

Table 1 | Sizes and vulnerabilities of belowground carbon stocks

Carbon 'pool'	Global estimates of size (Pg)	Potential loss by 2100 due to global warming (Pg)	References
Upland soil inventory (3 m depth)	2,300	–	3,4
Simulated upland soil (litter layer)	200	30	38
Simulated upland soil (mineral layer to 1 m depth)			
Annually cycling	20	3	38
Decadally cycling	700	40	38
Millennially cycling	100	0	38
Peatlands (3 m depth)	400–500	100	4
Permafrost (3 m depth)	400	100	4

Although the estimates here are highly uncertain (all estimates are rounded to one significant figure), they help to frame the debate about the relative importance of each type of belowground carbon as a potential feedback to climate change over the next few decades.

¹The Woods Hole Research Center, PO Box 296, Woods Hole, Massachusetts 02543, USA. ²Department of Biology, University of Antwerpen, Universiteitsplein 1, B-2610, Wilrijk, Belgium.

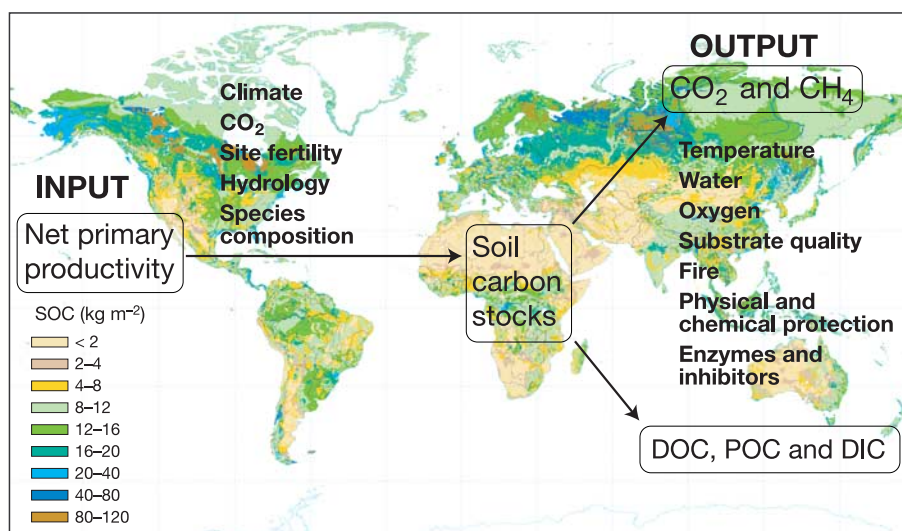
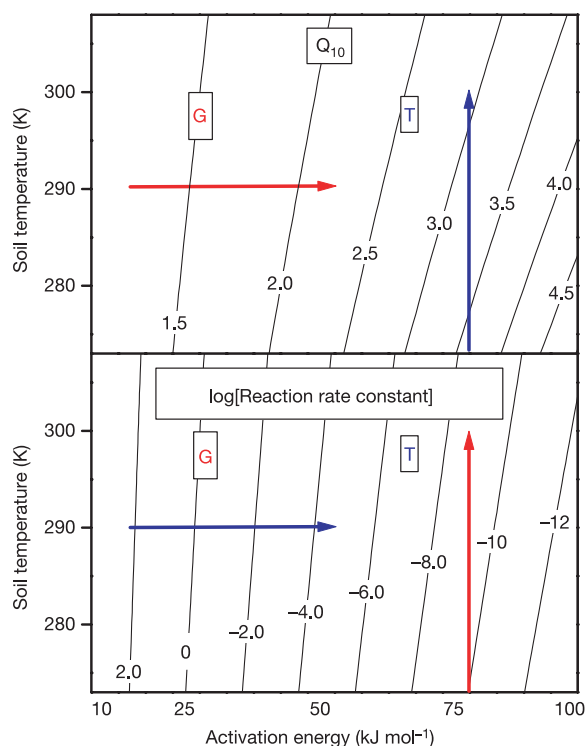


Figure 1 | Diagram of factors controlling the main inputs and outputs of soil carbon, superimposed over a global map of soil organic carbon stocks. While CO₂ is the main product of decomposition in soil, CH₄, dissolved organic carbon (DOC), particulate organic carbon (POC) in water, and dissolved inorganic carbon (DIC) are also significant exports from some soils. The background soil organic carbon (SOC) map (Miller projection; 1:100,000,000) is from ref. 100.

Box 1 | Must the Q_{10} of decomposition equal 2?

The Q_{10} for a reaction rate is defined as the factor by which the rate increases with a 10° rise in temperature. A rule of thumb widely accepted in the biological research community is that the rate of decomposition of SOM, like any other biological reaction rate, tends to double for every 10° rise in temperature (that is, the Q_{10} of decomposition is two). The origin of this rule-of-thumb, however, and the limits to its validity are less well known. Early experiments by van 't Hoff and colleagues indicated that, around room temperature, reaction rates "roughly double or triple" for every 10° rise in temperature (that is, reaction rates have Q_{10} values of the order of two to three)⁸⁴. Thus, historically, there has never been any suggestion that the Q_{10} should equal two on the basis of first principles. Moreover, an exponential equation does not always provide the desired relationship between the reaction rate and the temperature. Arrhenius noticed that a Q_{10} as high as two or three cannot originate from the increasing frequency of collisions between the reacting molecules, which only increases by about 1.5% for every 10° rise in temperature. Arrhenius also noticed that chemical reactions, even exergonic ones, often require a little 'push' to proceed, which he called the "activation energy" (E_a). He concluded that the explanation for the unexpected high temperature sensitivity of reaction rates had to be found in the amounts of reactants that possessed sufficient energy to react. Although the actual concentration of a reactant may be relatively constant with temperature, the active fraction that actually takes part in the reaction increases rapidly with temperature. Thus, Arrhenius developed the following equation⁸⁵: $k = a \exp(-E_a/RT)$ where k is the reaction rate constant; a is a frequency or pre-exponential factor (that is, the theoretical reaction rate constant in the absence of activation energy); E_a is the required activation energy; R is the gas constant ($8.314 \text{ J K}^{-1} \text{ mol}^{-1}$); and T is the temperature in Kelvin. The term $\exp(-E_a/RT)$ determines for any given temperature the fraction of the molecules present with energies equal to or in excess of the required activation energy. The Arrhenius function reveals some important properties of the reactions that it describes. For reactants with an E_a around 50 kJ mol^{-1} , and at temperatures between 273 K and 303 K, the Q_{10} of a chemical reaction is around two (Box 1 Figure, upper panel), in agreement with the rule-of-thumb cited above. However, the Arrhenius equation also predicts that the Q_{10} of chemical reactions decreases with increasing temperature (in Box 1 Figure, upper panel), as is also commonly observed in nature⁸⁶. The theoretical explanation for the decrease in Q_{10} with increasing temperature is that as temperature increases, there is a declining relative increase in the fraction of molecules with sufficient energy to react. The Arrhenius function also shows that reactants with higher activation energies (that is,

less reactive and more recalcitrant) should have higher temperature sensitivities (Box 1 Figure, upper panel). Hence, theoretical and experimental evidence shows that the Q_{10} of decomposition equals two only under specific conditions.



Box 1 Figure | Surface plots of the effects of temperature and activation energy on Q_{10} (upper panel) and reaction rate constant (bottom panel). Reaction rate constants are presented relative to that of glucose at room temperature. For clarity, the logarithm of this value is indicated. The red arrows indicate increasing trends and the blue arrows indicate decreasing trends. That the arrows switch colour between the two panels indicates that both temperature and activation energy have opposite effects on Q_{10} and on the reaction rate. The letters represent glucose (G) and tannin (T) in these graphic presentations of the Q_{10} and reaction rate constant values at room temperature.

temperature sensitivity of decomposition increases with increasing molecular complexity of the substrate (Boxes 1 and 2). The reaction rates are also modified by substrate concentrations and affinities of the enzymes for the substrates (Box 3).

In applying this knowledge to soil environments the controversy begins (hereafter we use 'soil' loosely to include wetlands, peatlands and permafrost). First, soils contain a 'veritable soup' of thousands of different organic-C compounds, each with its own inherent kinetic properties. Not only do plants produce a wide range of carbon substrates, but plant detritus also undergoes transformations by microbial degradation or by abiotic condensation reactions that produce new aromatic structures, larger molecular weights, insolubility, or other molecular architectures that affect the types and efficacies of enzymes that can degrade them^{25,26}. These complex molecular attributes are characterized by low decomposition rates, high activation energies, and inherently high temperature sensitivity (Box 1). We shall call the inherent kinetic properties based on molecular structure and ambient temperature the 'intrinsic temperature sensitivity' of decomposition.

Second, the enzymes for decomposition may be physically or chemically excluded from many of the organic-C substrates within the heterogeneous soil environment²⁶, causing substrate limitation at reaction microsites (Box 3). The observed response to temperature under these environmental constraints, which we shall call the 'apparent temperature sensitivity', may be much lower than the intrinsic temperature sensitivity of the substrate (Fig. 2). Conversely, if a temperature-sensitive process alleviates an environmental constraint to decomposition, then the subsequent increase in substrate availability could result in the apparent temperature sensitivity temporarily exceeding the intrinsic temperature sensitivity of the substrate.

The environmental constraints that can temporarily or indefinitely affect apparent temperature sensitivities of decomposition include the following:

Box 2 | Relative versus absolute changes in decomposition rates

The Arrhenius equation describes changes in relative reaction rates as a function of temperature, but the change in the absolute rate of the reaction is what concerns us most. While the relative rate of decomposition of recalcitrant soil organic matter with high activation energy may be very sensitive to temperature, the change in absolute rate may be small and difficult to detect in experiments. For example, in a temperate ecosystem and under current climate conditions, the annual decomposition of glucose (an easily degradable compound with an E_a of about 30 kJ mol⁻¹; ref. 87) would proceed 6.5 million times faster than annual decomposition of a tannin compound with an E_a of about 70 kJ mol⁻¹ (assuming equal and unlimited pool sizes; Box 1 Figure, lower panel). If the temperature were to increase by two degrees over the year, glucose decomposition would accelerate by 10%. Because of its higher temperature sensitivity, decomposition of the more recalcitrant tannin would accelerate by 21%. In absolute numbers, this difference in temperature response appears trivial, because glucose would still decompose 5.8 million times faster than the more recalcitrant tannin. Moreover, at similar levels of substrate availability, the warming-induced increase in C losses from glucose will be much larger than those associated with the decomposition of tannin, despite the lower Q_{10} of glucose decomposition. Hence, it is difficult to determine temperature responses of decomposition of recalcitrant organic matter in the presence of more labile compounds, and it is tempting to classify the temperature response of recalcitrant compounds as irrelevant. However, in most environments the stocks of labile and recalcitrant compounds are not equal, with recalcitrant compounds being much more abundant than easily degradable compounds. Thus, even a small change in their decomposition rate could become significant, albeit only at decadal or longer timescales. Hence, sensitivity to temperature changes must be evaluated within the context of pre-existing decomposition rates and substrate availability.

(1) Physical protection. Organic matter may become physically protected in the interior of soil aggregates^{27,28}, where microorganisms and their enzymes may have limited access and where oxygen concentrations may also be low. Similarly, organic compounds can be physically protected from degradation by water-soluble enzymes if they have low water solubility or if they occur in hydrophobic domains of humified organic matter²⁹.

(2) Chemical protection. Organic matter may become adsorbed onto mineral surfaces through covalent or electrostatic bonds, thus chemically protecting it from decomposition²⁷.

(3) Drought. Drought reduces the thickness of soil water films, thus inhibiting diffusion of extracellular enzymes and soluble organic-C substrates and lowering substrate availability at reaction microsites. In fire-prone or drought-prone regions, deposition of volatilized hydrophobic molecules can create water-repellency³⁰, which also restricts diffusion of organic matter and enzymes in water films.

(4) Flooding. Flooding slows oxygen diffusion to decomposition reaction sites, often allowing only anaerobic decomposition, which includes fewer and generally slower degradative enzymatic pathways.

(5) Freezing. Although enzymatic reactions can occur below 0 °C (refs 31, 32), the diffusion of substrates and extracellular enzymes within the soil is extremely slow where the extracellular soil water is frozen.

Each of these environmental constraints affects decomposition reaction rates, directly or indirectly, by decreasing substrate concentrations at enzymatic reaction sites. Instead of viewing decomposition

Box 3 | The effect of substrate availability on Q_{10}

The applicability of Arrhenius kinetics may be limited under conditions of changing substrate availability. The importance of substrate availability can easily be demonstrated in models of enzyme-catalysed processes. Enzymes affect reaction rates primarily by decreasing the required activation energy, such that they can occur at ambient temperatures. The importance of substrate availability in enzyme-catalysed reactions is described by Michaelis-Menten kinetics⁸⁸: the reaction rate is $V_{\max} \times [S]/(K_m + [S])$ where $[S]$ is the substrate availability (that is the substrate concentration at the active site of the enzyme), $V_{\max}$ is the maximum reaction rate at a given temperature, and K_m is the Michaelis-Menten constant, representing the substrate concentration at which the reaction rate equals $V_{\max}/2$. When $[S]$ is abundant, K_m becomes insignificant, and the temperature response of $V_{\max}$ determines that of the reaction rate. $V_{\max}$ increases with temperature^{89,90} up to an optimum temperature (typically well above ambient conditions), beyond which the enzyme starts to denature and $V_{\max}$ declines rapidly. Therefore, when $[S]$ is abundant and the temperature does not exceed the optimum temperature, $V_{\max}$ follows Arrhenius kinetics and the theory explained in Box 1 is valid. However, when $[S]$ is low, K_m becomes relevant. Because the K_m of most enzymes increases with temperature, the temperature sensitivities of K_m and $V_{\max}$ can neutralize each other, creating very low apparent Q_{10} values^{90,91}. Thus, in addition to substrate quality and temperature, temporal and spatial differences in substrate availability can also contribute to the large variability in Q_{10} observed in nature. This effect of substrate limitation on the temperature sensitivity of decomposition has not yet been incorporated into carbon cycle models. Even when combined, Arrhenius and Michaelis-Menten kinetics may not always describe decomposition reactions. First, these kinetics assume constant enzyme concentrations, which may not be the case when microbial populations fluctuate. Second, temperature can also affect enzyme conformation⁹², isozyme production (isozymes have similar functions but different temperature responses)^{93,94}, and microbial community structure (and thus enzymatic spectrum)⁵⁴. Changes in enzymatic properties, commonly referred to as 'temperature acclimation', could offset temperature-induced increases in respiratory activity. However, although the existence of these processes is beyond doubt, their ecological importance remains to be tested⁹⁵.

rates only through the lens of the temperature-dependent Arrhenius function of the maximum enzymatic reaction rate ($V_{\max}$; Box 1), the substrate concentration and the affinity of the enzyme for the substrate (k_M , as in Michaelis–Menten kinetics; Box 3) are also crucial for understanding the reaction rate and its sensitivity to temperature. Indeed, conditions of low substrate concentrations at active sites of enzymes may be the rule rather than the exception within the soil matrix.

Common approaches to modelling decomposition

Most efforts to characterize the kinetics of SOM decomposition have stratified carbon compounds into ‘pools’ that share similar mean residence times (MRTs) within the soil. The MRT is the inverse of the decomposition reaction rate (k) and therefore reflects a combination of inherent reactivity of the compound and the environmental constraints on its decomposition. The two best-known biogeochemical models of soil carbon dynamics—the CENTURY³³ and ROTH-C³⁴ models—compartmentalize soil carbon into 5–7 conceptual pools, including 2–4 pools of decomposable plant material near the soil surface (litter layer) and three pools of carbon in the mineral soil, with MRTs ranging from years to millennia (Fig. 3). Decomposition of the plant detritus in the litter layers is based on well-supported functions of climate and indices of substrate decomposability, such as carbon-to-nitrogen ratios and lignin content^{35,36}. The three C pools in the mineral soil, from the most labile to the most recalcitrant to decomposition, are called ‘fast’, ‘slow’ and ‘passive’ in CENTURY and ‘microbial biomass’, ‘humified organic matter’ and ‘inert’ in ROTH-C. Many attempts have been made with partial success to measure these various pools through physical and chemical fractionation of the soil^{13,28}, but they remain largely simplified modelling constructs. In lieu of discrete pools, a continuum of soil C substrates of varying chemical complexity and MRTs has also been used to simulate soil C dynamics³⁷.

Although direct measurements of the sizes and MRTs of these

conceptual pools of soil C remain imperfect, a consensus has emerged that using multi-pool soil C models to simulate changes in soil C stocks is a major improvement over treating soil C as a single, homogeneous pool^{38–40}. A substantial fraction of the SOM resides in the most recalcitrant pool that decomposes very slowly. The importance of this model structure was demonstrated when the multipool ROTH-C model was used in lieu of a single soil C pool model for a global simulation of climate change using the Hadley general circulation climate model. Soil C losses and gains were less severe with the multipool model, both regionally and globally³⁸.

While these models have proven effective for explaining local and regional variation in current soil C stocks and changes in stocks due to management and land-use change, a consensus has not emerged for their applicability to climate change. Typically, most models of soil C dynamics assume that decomposition of all SOM is nearly equally sensitive to temperature^{12,41,42} but this assumption is contrary to kinetic theory (ref. 43; Box 1). Moreover, decomposition rates may be slow (and MRTs may be long) either because the complex structures of the molecules render them resistant to decomposition, or because environmental constraints restrict access of enzymes to the molecules, or because of a combination of these two factors. Both protected simple compounds and more complex unprotected compounds might be lumped together into a common pool with common MRTs. If the causes of varying MRTs and their potential for change are to be understood, the distinction between intrinsic and apparent temperature sensitivities needs to be addressed explicitly.

Evidence for a decomposition feedback to warming

Discussions of biospheric feedbacks to climatic disruption have been influenced by the perspective that temperature is the dominant limiting factor of respiration, whereas photosynthesis is limited by multiple factors, including light, CO₂ concentration, water stress, and nutrient availability⁴⁴. Woodwell⁴⁵ and Jenkinson *et al.*⁴⁶ argued that respiration of terrestrial ecosystems, including microbial decomposition of SOM, would be more sensitive to global warming than would gross primary productivity. Global warming would thus lead to a net increase of C release to the atmosphere by the terrestrial biosphere (or less net C uptake from the atmosphere by the terrestrial biosphere). However, in the absence of a consensus on the temperature sensitivity of decomposition of a large fraction of soil C stocks, the significance of this positive feedback continues to be debated.

Current geographical relationships between climate and SOM stocks provide important clues, such as the presence of large soil C

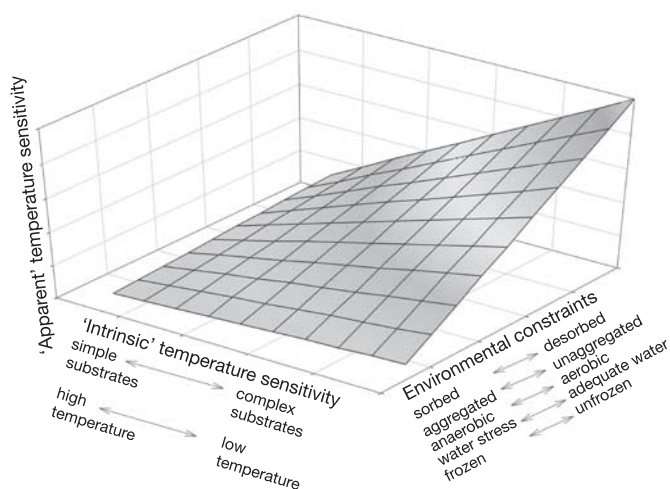


Figure 2 | Factors affecting the ‘apparent’ sensitivity of decomposition of soil organic matter. The intrinsic temperature sensitivity (as in Arrhenius functions, Box 1) of decomposition of an organic-C substrate is a function of the decomposability of the molecule and the ambient temperature. In general, more complex molecular structures have higher activation energies and, hence, higher temperature sensitivity. However, several environmental constraints on decomposition can dampen or obscure the intrinsic temperature sensitivity by reducing substrate availability, often causing the measured (or ‘apparent’) temperature sensitivity to be less than expected. The responses to these multiple factors are shown here as a plane for graphic simplicity, but nonlinear functions (for example, oxygen availability) or step functions (for example, melting of permafrost) may be more realistic for some factors.

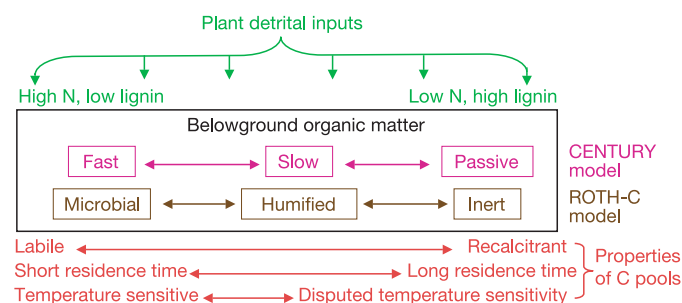


Figure 3 | Diagram of properties of conceptual pools of belowground carbon stocks in two well-known models. The CENTURY³³ and ROTH-C³⁴ models each define three discrete soil carbon pools in the mineral soil that lie roughly along a continuum of decomposability and MRT in the soil. The temperature sensitivity of decomposition of the more recalcitrant forms of C is the subject of recent debate. Much of this confusion is due to the fact that the recalcitrant pools are mixtures of simple compounds that have long MRTs owing to physical or chemical protection from decomposition and more complex compounds that have inherently low reactivity and require high activation energy for decomposition.

stocks at high latitudes (Fig. 1). Inferences about the effects of changing climate based on current geographic patterns, however, require the tenuous assumption that a space-for-time substitution adequately describes projections of future scenarios of changing carbon stocks. Nevertheless, in his classical work on soil-forming factors, Jenny^{47,48} described how soil N concentrations (which covary with soil C) increased with decreasing temperature and increasing precipitation across the Great Plains of central North America. Kirschbaum⁴⁹ pointed out that Jenny's results imply that decomposition does indeed increase more with temperature than does net primary productivity across this gradient, supporting Woodwell's hypothesis⁴⁵. Post *et al.*⁵⁰ demonstrated similar trends globally, although strong relationships were apparent only for very wet and very dry climates, suggesting that soil water content also plays an important role for constraining rates of decomposition.

Not all gradient studies have supported temperature sensitivity of decomposition¹¹, and there are numerous possible explanations. Notwithstanding Jenny's example^{47,48} it is possible that growing-season length along some north-south gradients could affect gross primary productivity, and thus C inputs to soil, as much or more than temperature affects respiration, thus obscuring the effect of temperature on decomposition when interpreting variation in SOM stocks⁵¹. Environmental constraints to decomposition could also covary with temperature along a particular study transect if factors such as mineralogy, clay content, aggregation, or soil water content also covaried along the same gradient. For example, the presence of clay-sized minerals that effectively adsorb organic matter and retain soil moisture can be related to temperature-dependent processes, such as the effects of previous glaciations that expose new bedrock and biogenic production of acids that promote mineral weathering⁵². Finally, in the example of an observed positive relationship between temperature and soil C stocks along a climate gradient in Finland¹¹, the inferred temperature-insensitivity of decomposition of a large fraction of SOM could be an artefact of model assumptions about fixed or variable mean residence times of conceptualized soil C pools⁵³. We will return to this issue of the consequences of model assumptions on interpreted temperature sensitivities.

When analysing any geographical trend, it is important to remember that instantaneous temperature responses of decomposition of current C stocks reflect the relative abundances of organic-C substrates of differing kinetic properties. Those relative substrate abundances result, in part, from environmental constraints to decomposition during the climate and disturbance history of the soil. For example, the decomposition of organic matter in a highly weathered mature tropical forest soil with high clay content may, on average, have low apparent temperature sensitivity because of chemical protection of a large fraction of soil C on mineral surfaces. In contrast, the apparent temperature sensitivity may, on average, be higher in a recently tilled temperate prairie soil because of lower proportions of substrates under environmental constraints to decomposition. Giardina and Ryan⁶ ignored this variation in relative abundances of different soil substrates when they calculated MRTs of a single homogeneous soil C pool for a number of soil samples from different latitudes and under different temperature regimes in laboratory experiments. Not surprisingly, they found no correlation between calculated total soil C MRTs and either laboratory incubation temperature or the mean annual ambient temperature of the field locations. They correctly proposed that substrate quality and other stabilization mechanisms are important factors affecting variation of soil C stocks among study sites, but they incorrectly concluded that the temperature sensitivity of decomposition is unimportant. Temperature insensitivity implies zero activation energies of decomposition, which is impossible for biochemical processes. The analysis by Giardina and Ryan⁶ also did not take into account that the abundance of the various carbon substrates (or 'pools'), with differing intrinsic temperature sensitivity and under

differing constraints to decomposition, are themselves partly the result of climate effects (Box 4).

In addition to observations of natural gradients, several studies of the temperature sensitivity of decomposition have been carried out in the laboratory and in field experiments. Fang *et al.*⁹ experimentally applied multiple cycles of varying temperature during a 108-day incubation and estimated the temperature sensitivity for each cycle. The temperature sensitivities of decomposition were not significantly different for the most labile carbon that was respired early in the incubation and for the less labile carbon being respired at the end of the incubation period. They also conducted the experiment on soils from the surface and from deeper soil horizons, the latter presumably having more recalcitrant C associated with mineral surfaces. Again, no statistically significant differences in temperature sensitivities were observed. While the results of Fang *et al.*⁹ refuted the notion proposed by others^{6,11,54,55} that decomposition of the more recalcitrant pools would be less sensitive to temperature than the more labile pools, their results are still contrary to kinetic theory, which indicates higher intrinsic temperature sensitivity for decomposition of the recalcitrant C pools.

In contrast, Knorr *et al.*¹⁰ came to a conclusion consistent with kinetic theory when they fitted data from a laboratory soil incubation experiment⁵⁶ to a multi-pool soil C model. They calculated not only that decomposition responded positively to temperature for a highly labile pool, but also that decomposition of a less labile pool exhibited higher temperature sensitivity. Most of the SOM resided in a third recalcitrant pool that did not decompose significantly during the

Box 4 | Assumptions of curve fitting

In the Arrhenius function shown in Box 1, E_a is clearly a variable that is fitted to observational data to determine a temperature sensitivity of a reaction. A debate has recently emerged, however, whether the pre-exponential factor (a), should be a constant or a variable^{57,61,62}. From a purely mathematical perspective, the statistical fits of E_a and a are not independent, and so the decision to keep a constant or allow it to vary affects the fitted value of E_a (ref. 62). Because the a -term is defined as the theoretical reaction rate constant in the absence of activation energy, it makes sense that it might have different values for different substrates⁵³. Hence, model fits in which all variation in decomposition rates and temperature sensitivity is attributed to E_a are likely to overemphasize the differences in E_a among compounds. Nonetheless, there is overwhelming evidence that E_a does vary considerably among compounds, and that decomposition of recalcitrant SOM has higher intrinsic temperature sensitivity than decomposition of labile SOM (Box 1). Because the statistical fits of E_a and a are not independent, however, curve fitting is unlikely to have sufficient power to resolve exactly to what degree decomposition of recalcitrant compounds should have higher intrinsic temperature sensitivities. A conceptually separate, but mathematically similar, issue arises when applying the rate constant (k) from the Arrhenius function to simulate decomposition of various soil C pools, using the form: $g(t) = \sum c_i e^{-k_i t}$, where $g(t)$ is the remaining carbon fitted to observational data and c_i are the initial sizes of carbon 'pools' of varying degrees of decomposability. Note the mathematical similarity between this function and the Arrhenius function (Box 1), from which the k_i value for each c_i in this equation is derived. Just as the variability of the a -term in the Arrhenius function is being debated, a debate has also emerged in the literature as to whether temperature dependence should reside only in the k_i -terms or also in the c_i -terms⁹⁶⁻⁹⁹. While the k_i -term describes the current instantaneous decomposition rate and its temperature sensitivity, the relative sizes of the carbon pools (c_i) of varying degrees of decomposability were determined over longer timescales and may also partially be a consequence of climatic history, including temperature. In many statistical fits of observational data, these two terms (k_i and c_i) are correlated⁹⁸, and so it is impossible to determine where the temperature sensitivity lies by statistical curve fitting alone.

incubation period. If not partitioned out in the model, this large pool of recalcitrant organic matter would obscure the temperature sensitivity of the two smaller, more labile pools. Experimental support for kinetic theory was also obtained from incubation experiments of leaf and root litter with varying substrate quality¹⁴.

However, each of these studies^{9–11,14} relied upon curve fitting to demonstrate temperature sensitivities to decomposition, and doubts may arise from the underlying assumptions of the models and the power of the statistical tests to detect significantly different curve-fitting parameters. It is possible, for example, that Fang *et al.*⁹ did not find significantly higher temperature sensitivity of the more recalcitrant C pools because of a type II error—accepting a null hypothesis of no statistically significant differences among fitted E_a terms. On the other hand, the Knorr *et al.*¹⁰ study has been criticized for creating an artefact of different temperature sensitivities (E_a terms) between pools because of a model assumption of fixed a -terms in the Arrhenius function (ref. 57; Boxes 1 and 4). Furthermore, environmental constraints to decomposition of some of the substrates may have been present during these incubations.

Soil warming experiments in the field have also provided equivocal evidence regarding the temperature sensitivity of decomposition. In these experiments, an initial increase in soil CO₂ efflux in response to experimental warming has been observed, but this measurable pulse in decomposition often disappears within a few years^{54,55,58,59}. This result, however, must be interpreted with caution because high spatial and temporal variability of measured CO₂ efflux rates often preclude powerful statistical tests of small treatment effects, resulting in possible type II errors. Nonetheless, one interpretation of these ephemeral responses has been that only decomposition of the most labile soil C pool was sensitive to the warming treatment, and that decomposition of the older, more recalcitrant soil C was not temperature sensitive. Another interpretation has been that roots and soil microbial communities ‘acclimate’ to the higher-temperature conditions by gradually adapting their metabolism, so that the newly acclimated communities return to respiration rates similar to the pre-treatment levels within a relatively short time (Box 3). However, recent modelling studies^{10,59,60} have demonstrated that these field experimental data are also consistent with a model of a small, labile pool that is quickly exhausted and a larger less-labile pool that is also temperature sensitive, but that decomposes much more slowly. Model fits of these studies appear better with the multipool models, and neither temperature insensitivity nor acclimation needs to be invoked to explain the observations.

Unfortunately, curve fitting of data from laboratory incubations^{6,9–11,14} and soil warming experiments in the field^{54,55,58–60} is unlikely to resolve the debate about the underlying mechanisms of temperature sensitivities. In a series of discussions of the Knorr *et al.*¹⁰ and Fang *et al.*⁹ papers, the same data fit equally well models with different assumptions and, hence, different interpretations of the temperature sensitivity of decomposition of recalcitrant SOM (refs 57, 61, 62; Box 4). Another limitation to curve fitting is that it focuses on the relationship between measured apparent temperature sensitivity and estimated MRTs of experimentally defined soil C pools, rather than attempting to distinguish between intrinsic and apparent temperature sensitivity of the substrates of decomposition.

Despite these controversies, the observational data are converging to demonstrate that more than one ‘pool’ along the labile/recalcitrant continuum (Fig. 3) decomposes with detectable apparent temperature sensitivity. Neither the observational data nor the soil C model simulations, however, have been able to demonstrate a consistent apparent temperature sensitivity of decomposition along the entire spectrum of recalcitrant pools of SOM.

It is tempting to assume that the temperature sensitivity of decomposition of the more recalcitrant forms of SOM is trivial with respect to current concern about feedbacks to global warming, because decomposition of these substrates, even if it is accelerated, contributes so little to instantaneous CO₂ fluxes. Nonetheless,

because of their large contribution to the soil C stocks, even small changes in the decomposition rates of recalcitrant pools may result in an important change in soil C stocks over decades (Box 2). In fact, we do not know where along the soil-C continuum or among the discrete conceptual pools of SOM (Fig. 3) apparent temperature sensitivity may become irrelevant to contemporary issues of carbon cycling. More importantly, might the existing environmental constraints to decomposition change with changing climate, thus exposing to decomposition SOM with high intrinsic temperature sensitivity? In our opinion, answering this question would be more informative about biospheric feedbacks to warming than refining the correlation between current apparent temperature sensitivity and MRTs of soil C pools.

Temperature dependence of environmental constraints

According to kinetic theory, the constraints to decomposition that are caused by biological and chemical processes must themselves be affected by temperature⁸ and perhaps other climatic drivers. Let us now reexamine the five environmental constraints to decomposition listed above in light of the question of their own dependence on climate.

- (1) Both climate and management affect aggregate formation (through growth of fungal hyphae and activity of soil fauna), which physically protects SOM. The breakdown of aggregates can also be enzymatic²⁶, as the biogenic ‘glue’ that holds the aggregates together is decomposed. In addition, however, purely physical processes, such as ploughing and the impact of raindrops, also destroy aggregates²⁶. These processes are not directly temperature dependent, but are often influenced by climate.
- (2) Temperature affects the chemical processes of SOM adsorption and desorption onto mineral surfaces, but little is known about the activation energies of these processes.
- (3) The climate-driven hydrologic balance among drainage, precipitation, and evapotranspiration determines soil water film thickness through which the diffusion of soluble organic-C substrates and extracellular enzymes occurs. Likewise leaf litter hydrophobicity associated with drought-prone and fire-prone ecosystems is also affected by climate.
- (4) Climate-driven flooding of wetlands and peatlands determines oxygen supply for decomposition. Both precipitation and evapotranspiration are likely to change in many regions of the world owing to climatic disruption⁶³, tipping the hydrologic balance towards summertime drying of many mid-continental peatlands and wetlands and thus exposing large stocks of carbon substrates to aerobic decomposition.
- (5) Melting of permafrost will expose organic matter with wide-ranging kinetic properties that are not currently expressed in frozen soil. Once the soil thaws, a major constraint to decomposition will be removed.

Most of the current debate regarding the temperature sensitivity of decomposition of organic matter in mineral soils concerns only physical and chemical protection within the mineral soil matrix. While these processes are extremely important contributors to variability of soil carbon stocks and soil fertility, the extent to which they will participate in positive or negative feedbacks to climate change is not clear. Adsorption and desorption processes are both temperature sensitive and might both increase such that the net effect would be minimal over the next few decades. In contrast, the last two constraints to decomposition—frozen soils and oxygen limitation due to flooding—are likely to be subject to rapid changes under plausible climate change scenarios.

When decomposition in wetlands and peatlands slows owing to lack of oxygen during periods of flooding, the low oxygen concentrations inhibit the activity of phenol oxidase, causing accumulation of phenolic compounds⁶⁴. These phenolic compounds inhibit the activity of hydrolase enzymes responsible for decomposition, thus slowing decomposition further. This inhibition is quickly reversible

once peat becomes aerobic. Hence, the 400–500 Pg (1 Pg = 10^{15} g) carbon in wetlands and peatlands^{4,65}, which has been accumulating over centuries and millennia, is ‘stable’ only as long as anaerobic conditions are sustained. In the continental areas where summertime soil moisture is expected to decrease⁶³, the upper layers of peat could dry out. An estimated 100 Pg carbon could become aerobic and thus available for decomposition (Table 1; ref. 4). Evidence that this process may already be occurring comes from recently repeated inventories of soils of England and Wales, which show that peat soils and bogs lost carbon at a faster rate than upland soils over the last 25 years (ref. 5).

Carbon losses from peatlands will not necessarily enhance global warming if an increased emission of CO₂ is compensated by a decrease in the current net emissions of CH₄ (ref. 66). The greenhouse warming potential of CH₄ on a per molecule basis is 23 times higher than CO₂ on a 100-year timescale⁶⁷. Likewise, growth of forest vegetation on previously flooded land could sequester significant amounts of C in wood⁶⁸, although the net carbon balance remains uncertain⁶⁹. On the other hand, carbon in desiccated peat is also subject to natural and human fires, rapidly releasing huge amounts of carbon to the atmosphere. Siberian and Canadian peatlands are already subject to important peat losses during fire-prone dry years⁷⁰ and the combination of higher temperatures and peat drying may increase fire frequency and severity⁷¹. During the 1997 El Niño, 0.6–0.8 Pg of C (10% of anthropogenic emissions) was lost owing to peat fires in Indonesia alone⁷². Peat fires could thus have larger effects on the soil carbon feedback than any of the biotic responses presented above⁷³.

Permafrost soils store a similar amount of organic matter as peatlands (Table 1). In these soils with permanently frozen layers, plant litter accumulates both at the surface and on top of the permafrost table through a mixing process called cryoturbation¹⁶. When permafrost thaws, large amounts of otherwise mostly unprotected carbon become available for decomposition^{74,75}. A gradual net loss of deep soil C in a boreal forest has also been attributed to warming-induced deepening of the layer of seasonal biological activity⁷⁶. One estimate suggests that global warming could thaw 25% of the permafrost area by 2100 (ref. 77), thus rendering about 100 Pg carbon vulnerable to decay (Table 1; ref. 4).

Permafrost thaw creates a mosaic of flooded areas interspersed within higher dry areas. In the drier thawed areas, much of the large substrate pool is likely to decompose relatively quickly. Within the flooded thawed areas (thermokarst lakes), however, anaerobic decomposition of organic matter is likely to proceed more slowly, but produces large CH₄ emissions⁷⁸, which could constitute a stronger feedback to the climate system than the larger soil C losses from the drier areas. To further complicate matters, recent evidence indicates that thermokarst lakes are increasing in abundance in the most northern range of permafrost, owing to an initial increase in thermokarst development in response to warming, whereas thermokarst lakes are draining and disappearing in the southern range of permafrost, owing to more advanced degradation of permafrost⁷⁹.

Frozen and anaerobic conditions merely suspend organic matter in decomposition time, rather than transform it into inherently recalcitrant material. The term ‘stabilization’ has not been rigorously defined with respect to soil carbon dynamics, and it would be misleading to refer to organic C in wetlands and permafrost as ‘stabilized’ in the same sense that organic C is stabilized in mineral soils when it is adsorbed to mineral surfaces or protected in aggregate interiors. In both cases, intrinsic kinetic properties of decomposition of the organic-C substrates are suppressed by environmental constraints (Fig. 2), but the anaerobic and frozen conditions of wetland, peatland, and permafrost soils are more likely to be subject to rapid change. Unlike the debate over mineral soils, the unresolved question regarding peatlands and permafrost is not the degree to which the currently constrained decomposition rates are temperature sensitive,

but rather how much permafrost is likely to melt and how much of the peatland area is likely to dry significantly. Such regional changes in temperature, precipitation, and drainage are still difficult to predict in global circulation models. Hence, the climate change predictions, as much as our understanding of carbon dynamics, limit our ability to predict the magnitude of likely vulnerability of peat and permafrost carbon to climate change. Assuming that 25% of the estimated stocks of carbon in peatlands and permafrost is subject to loss due to global warming in the twenty-first century, this potential loss would be two to three times larger than simulated C losses from mineral soils using current soil C models (Table 1; refs 38, 42).

Other global change processes, such as CO₂ fertilization, N deposition, improved soil management (for example, conservation tillage), and land-use change could also change soil carbon stocks. We have not addressed these changes, partly because our charge in this review is to focus on the temperature sensitivity controversy, but also because compelling evidence is lacking for globally significant soil C sinks by these processes^{15,80–83}. In contrast, those belowground carbon pools where environmental constraints to decomposition are themselves highly sensitive to climate may become increasingly important positive feedbacks as global climatic disruption becomes more pronounced.

Conclusions and future research directions

A significant fraction of relatively labile SOM is clearly subject to temperature-sensitive decomposition, but another significant fraction of SOM remains under environmental constraints that often obscure the intrinsic temperature sensitivity of its decomposition. The interpretations of natural climatic gradients and of laboratory and field experiments designed to quantify various carbon fractions and degrees of temperature sensitivity are highly dependent upon model assumptions and curve fitting techniques. Such studies have yielded valuable insight into soil carbon dynamics, but they have not resolved the overall response of global soil C stocks or the magnitude of expected feedbacks to climatic disruption. Moreover, dividing SOM into only temperature-sensitive and apparently temperature-insensitive pools is far too simplistic.

Extrapolation of decomposition rates into a future warmer world based on observations of current apparent temperature sensitivities is inadequate. Rather, we need to understand how substrate availability will change and how a changing set of environmental constraints to decomposition in a future climate will determine the future apparent temperature sensitivity of decomposition. Perhaps a new way forward would be a further reductionist effort to distinguish between inherent kinetic properties of individual substrates and the suite of environmental constraints to decomposition that frequently exist *in situ*. Our ability to identify and characterize soil substrates is growing with new developments in nuclear magnetic resonance and other technologies, but the task is enormous given the diversity of soil C substrates common in soils. Even when structures are identified within the soil, their concentrations at reactive sites of enzymes are more difficult to estimate. Nevertheless, merging the concepts of substrate availability, such as Michaelis–Menten kinetics, with the temperature sensitivity prescribed by Arrhenius kinetics may provoke new measurement and modelling approaches for soil C dynamics. The multiple processes of environmental constraints that govern availability of substrates to enzymes should be explicitly described and studied within the context of climate change.

Regardless of the experimental and modelling approaches used, the debate about the temperature sensitivity of decomposition should be broadened beyond upland mineral soils specifically to include wetlands, peatlands and permafrost soils. These are the most obvious environments in which current constraints on decomposition are likely to change as a result of climatic disruption, thus potentially exposing large stocks of C to less constrained decomposition during the next few decades. A high research priority should be

how the constraints to decomposition in these environments are sensitive to climate.

1. Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **47**, 151–163 (1996).
2. Eswaran, H., Van Den Berg, E. & Reich, P. Organic carbon in soils of the world. *Soil Sci. Soc. Am. J.* **57**, 192–194 (1993).
3. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
4. Gruber, N. *et al.* in *Toward CO₂ Stabilization: Issues, Strategies, and Consequences* (eds Field, C. B. & Raupach, M. R.) 45–76 (Island Press, Washington, DC, 2004).
5. Bellamy, P. H., Loveland, P. J., Bradley, R. I., Murray Lark, R. & Kirk, G. Carbon losses from all soils across England and Wales 1978–2003. *Nature* **437**, 245–248 (2005).
6. Giardina, C. P. & Ryan, M. G. Evidence that decomposition rates of organic matter in mineral soil do not vary with temperature. *Nature* **404**, 858–861 (2000).
7. Davidson, E. A., Trumbore, S. E. & Amundson, R. Soil warming and organic carbon content. *Nature* **408**, 789–790 (2000).
8. Thornley, J. H. M. & Cannell, M. G. R. Soil carbon storage response to temperature: an hypothesis. *Ann. Bot. (Lond.)* **87**, 591–598 (2001).
9. Fang, C., Smith, P., Moncrieff, J. B. & Smith, J. U. Similar response of labile and resistant soil organic matter pools to changes in temperature. *Nature* **433**, 57–59 (2005).
10. Knorr, W., Prentice, I. C., House, J. I. & Holland, E. A. Long-term sensitivity of soil carbon turnover to warming. *Nature* **433**, 298–301 (2005).
11. Liski, J., Ilvesniemi, H., Mäkelä, A. & Westman, C. J. CO₂ emissions from soil in response to climatic warming are overestimated— the decomposition of old soil organic matter is tolerant of temperature. *Ambio* **28**, 171–174 (1999).
12. Kirschbaum, M. U. F. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biol. Biochem.* **27**, 753–760 (1995).
13. Trumbore, S. E., Chadwick, O. A. & Amundson, R. Rapid exchange between soil carbon and atmospheric carbon dioxide driven by temperature change. *Science* **272**, 393–396 (1996).
14. Fierer, N., Craine, J. M., McLaughlin, K. & Schimel, J. P. Litter quality and the temperature sensitivity of decomposition. *Ecology* **86**, 320–326 (2005).
15. Schlesinger, W. H. & Andrews, J. A. Soil respiration and the global carbon cycle. *Biogeochemistry* **48**, 7–20 (2000).
16. Bockheim, J. G. & Tarnocai, C. Recognition of cryoturbation for classifying permafrost-affected soils. *Geoderma* **81**, 281–293 (1998).
17. Arctic Climate Impact Assessment (ACIA). *Impacts of a Warming Arctic: Arctic Climate Impact Assessment* (Cambridge Univ. Press, Cambridge, 2004); (<http://amap.no/acia/>).
18. Chapin, F. S. III *et al.* Role of land-surface changes in Arctic summer warming. *Science* **310**, 657–660 (2005).
19. Davidson, E. A., Verchot, L. V., Cattaneo, J. H., Ackerman, I. L. & Carvalho, J. E. M. Effects of soil water content on soil respiration in forests and cattle pastures of eastern Amazonia. *Biogeochemistry* **48**, 53–69 (2000).
20. Hanson, P. J. *et al.* in *North American Temperate Deciduous Forest Responses to Changing Precipitation Regimes* (eds Hanson, P. J. & Wulfschleger, S. D.) 163–189 (Springer, New York, 2003).
21. Janssens, I. A. *et al.* in *Canopy Fluxes of Energy, Water and Carbon Dioxide of European Forests* (ed. Valentini, R.) 235–256 (Springer, Berlin, 2003).
22. Raich, J. W. & Potter, C. S. Global patterns of carbon dioxide emissions from soils. *Glob. Biogeochem. Cycles* **9**, 23–36 (1995).
23. Reichstein, M. *et al.* Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. *Glob. Biogeochem. Cycles* **17**, 15.1–15.15 (2003).
24. Schlentner, R. E. & Van Cleve, K. Relationships between CO₂ evolution from soil, substrate temperature, and substrate moisture in four mature forest types in interior Alaska. *Can. J. For. Res.* **15**, 97–106 (1985).
25. Kelley, K. R. & Stevenson, F. J. in *Humic Substances in Terrestrial Ecosystems* (ed. Picolo, A.) 407–427 (Elsevier, Amsterdam, 1996).
26. Sollins, P., Homann, P. & Caldwell, B. A. Stabilization and destabilization of soil organic matter: mechanisms and controls. *Geoderma* **74**, 65–105 (1996).
27. Oades, J. M. The retention of organic matter in soils. *Biogeochemistry* **5**, 35–70 (1988).
28. Six, J. *et al.* Measuring and understanding carbon storage in afforested soils by physical fractionation. *Soil Sci. Soc. Am. J.* **66**, 1981–1987 (2002).
29. Spacini, R. *et al.* Increased soil organic carbon sequestration through hydrophobic protection by humic substances. *Soil Biol. Biochem.* **34**, 1839–1851 (2002).
30. McHale, G. *et al.* Water-repellent soil and its relationship to granularity, surface roughness and hydrophobicity: a materials science view. *Eur. J. Soil Sci.* **56**, 445–452 (2005).
31. Mikan, C. J. *et al.* Temperature controls of microbial respiration in arctic tundra soils above and below freezing. *Soil Biol. Biochem.* **34**, 1785–1795 (2002).
32. Elberling, B. & Brandt, K. K. Uncoupling of microbial CO₂ production and release in frozen soil and its implications for field studies of arctic C cycling. *Soil Biol. Biochem.* **35**, 263–272 (2003).
33. Parton, W. J., Schimel, D. S., Cole, C. V. & Ojima, D. S. Analysis of factors controlling soil organic matter levels in Great Plains Grasslands. *Soil Sci. Soc. Am. J.* **51**, 1173–1179 (1987).
34. Jenkinson, D. S. The turnover of organic carbon and nitrogen in soil. *Phil. Trans. R. Soc. B* **329**, 361–368 (1990).
35. Meentemeyer, V. Macroclimate and lignin control of litter decomposition rates. *Ecology* **59**, 465–472 (1978).
36. Melillo, J. M. *et al.* Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. *Ecology* **63**, 621–626 (1982).
37. Ågren, G. I. & Bosatta, E. *Theoretical Ecosystem Ecology—Understanding Element Cycles* (Cambridge Univ. Press, Cambridge, 1998).
38. Jones, C. *et al.* Global climate change and soil carbon stocks; predictions from two contrasting models for the turnover of organic carbon in soil. *Glob. Change Biol.* **11**, 154–166 (2005).
39. Powlson, D. Will soil amplify climate change? *Nature* **433**, 204–205 (2005).
40. Trumbore, S. Age of soil organic matter and soil respiration: radiocarbon constraints on belowground C dynamics. *Ecol. Appl.* **10**, 399–411 (2000).
41. Jenkinson, D. S. & Rayner, J. H. The turnover of soil organic matter in some of the Rothamsted Classical Experiments. *Soil Sci.* **123**, 298–305 (1977).
42. Schimel, D. S. *et al.* Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Glob. Biogeochem. Cycles* **8**, 279–293 (1994).
43. Bosatta, E. & Ågren, G. I. Soil organic matter quality interpreted thermodynamically. *Soil Biol. Biochem.* **31**, 1889–1891 (1999).
44. Farquhar, G. D., von Caemmerer, S. & Berry, J. A. A biochemical model of photosynthetic CO₂ fixation in leaves of C₃ species. *Planta* **149**, 78–90 (1980).
45. Woodwell, G. M. in *Changing Climate* 216–241 (NAS Press, Washington DC, 1983).
46. Jenkinson, D. S. *et al.* model estimates of CO₂ emissions from soil in response to global warming. *Nature* **351**, 304–306 (1991).
47. Jenny, H. *Factors of Soil Formation. A System of Quantitative Pedology* 1st edn (McGraw-Hill Book Co, New York, 1941).
48. Jenny, H. *The Soil Resource, Origin and Behavior* (Springer, New York, 1980).
49. Kirschbaum, M. U. F. Will changes in soil organic carbon act as a positive or negative feedback on global warming. *Biogeochemistry* **48**, 21–51 (2000).
50. Post, W. M., Emanuel, W. R., Zinke, P. J. & Stangenberger, A. G. Soil carbon pools and world life zones. *Nature* **298**, 156–159 (1982).
51. Janssens, I. A. *et al.* Productivity overshadows temperature in determining soil and ecosystem respiration across European forests. *Glob. Change Biol.* **7**, 269–278 (2001).
52. Chadwick, O. A. *et al.* Carbon dioxide consumption during soil development. *Biogeochemistry* **24**, 115–127 (1994).
53. Ågren, G. Temperature dependence of old soil organic matter. *Ambio* **29**, 55 (2000).
54. Luo, Y., Wan, S. & Hui, D. Acclimatization of soil respiration to warming in tall grass prairie. *Nature* **413**, 622–625 (2001).
55. Melillo, J. M. *et al.* Soil warming and carbon-cycle feedbacks to the climate system. *Science* **298**, 2173–2175 (2002).
56. Holland, E. A. *et al.* Uncertainties in the temperature sensitivity of decomposition in tropical and subtropical ecosystems: Implications for models. *Glob. Biogeochem. Cycles* **14**, 1137–1151 (2000).
57. Fang, C. *et al.* Is resistant soil organic matter more sensitive to temperature than the labile organic matter? *Biogeosci. Discuss.* **2**, 725–735 (2005).
58. Rustad, L. E. *et al.* A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming. *Oecologia* **126**, 543–562 (2001).
59. Eliasson, P. E. *et al.* The response of heterotrophic CO₂-flux to soil warming. *Glob. Change Biol.* **11**, 167–181 (2005).
60. Kirschbaum, M. U. F. Soil respiration under prolonged soil warming: are rate reductions caused by acclimation or substrate loss? *Glob. Change Biol.* **10**, 1870–1877 (2004).
61. Knorr, W. *et al.* On the available evidence for the temperature dependence of soil organic carbon. *Biogeosci. Discuss.* **2**, 749–755 (2005).
62. Reichstein, M. *et al.* Does the temperature sensitivity of decomposition vary with soil organic matter quality? *Biogeosci. Discuss.* **2**, 737–747 (2005).
63. Wetherald, R. T. & Manabe, S. Detectability of summer dryness caused by greenhouse warming. *Clim. Change* **43**, 495–511 (1999).
64. Freeman, C., Ostle, N. & Kang, H. An enzymatic 'latch' on a global carbon store. *Nature* **409**, 149 (2001).
65. Maltby, E. & Immirzi, P. Carbon dynamics in peatlands and other wetland soils, regional and global perspectives. *Chemosphere* **27**, 999–1023 (1993).
66. Whalen, S. C., Reeburgh, W. C. & Reeburgh, W. S. Consumption of atmospheric methane by tundra soils. *Nature* **346**, 160–162 (1990).
67. Ramaswamy, V. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 349–416 (Cambridge Univ. Press, Cambridge, UK, 2001).
68. Shaver, G. S. *et al.* Global change and the carbon balance of arctic ecosystems. *Bioscience* **42**, 433–441 (1992).
69. Grogan, P. & Jonasson, S. Temperature and substrate controls on intra-annual variation in ecosystem respiration in two subarctic vegetation types. *Glob. Change Biol.* **11**, 465–475 (2005).
70. Turetsky, M. R., Wieder, R. K., Halsey, L. A. & Vitt, D. H. Current disturbance and the diminishing peatland carbon sink. *Geophys. Res. Lett.* **29**, 10.1029/2001GL014000 (2002).

71. Turetsky, M. R., Amiro, B. D., Bosch, E. & Bhatti, J. S. Historical burn area in western Canadian peatlands and its relationship to fire weather indices. *Glob. Biogeochem. Cycles* **18**, 4014, doi:10.1029/2004GB002222 (2004).
72. Page, S. E. *et al.* The amount of carbon released during peat and forest fires in Indonesia during 1997. *Nature* **420**, 61–65 (2002).
73. Chapin, F. S. III *et al.* Arctic and boreal ecosystems of western North America as components of the climate system. *Glob. Change Biol.* **6** (suppl. 1), 211–233 (2000).
74. Osterkamp, T. E. & Romanovsky, V. E. Evidence for warming and thawing of discontinuous permafrost in Alaska. *Permafrost Periglacial Process.* **10**, 17–37 (1999).
75. Serreze, M. C. *et al.* Observational evidence of recent change in the northern high-latitude environment. *Clim. Change* **46**, 159–207 (2000).
76. Goulden, M. L. *et al.* Sensitivity of boreal forest carbon balance to soil thaw. *Science* **279**, 214–217 (1998).
77. Anisimov, O. A., Nelson, F. E. & Pavlov, A. V. Predictive scenarios of permafrost development under conditions of global climate change in the XXI century. *Earth Cryology* **3**, 15–25 (1999).
78. Zimov, S. A. *et al.* North Siberian lakes: a methane source fueled by Pleistocene carbon. *Science* **277**, 800–802 (1997).
79. Smith, L. C., Sheng, Y., MacDonald, G. M. & Hinzman, L. D. Disappearing arctic lakes. *Science* **308**, 1429 (2005).
80. Ogle, S. M., Breidt, F. J., Eve, M. D. & Paustian, K. Uncertainty in estimating land use and management impacts on soil organic carbon storage for US agricultural lands between 1982–1997. *Glob. Change Biol.* **9**, 1521–1542 (2003).
81. Post, W. M. *et al.* Enhancement of carbon sequestration in US soils. *Bioscience* **54**, 895–908 (2004).
82. Telles, E. C. C. *et al.* Influence of soil texture on carbon dynamics and storage potential in tropical forest soils of Amazonia. *Glob. Biogeochem. Cycles* **17**, 1040, 9–1–12, doi:10.1029/2002GB001953 (2003).
83. Lichter, J. *et al.* Soil carbon sequestration and turnover in a pine forest after six years of atmospheric CO₂ enrichment. *Ecology* **86**, 1835–1847 (2005).
84. van 't Hoff, J. H. *Lectures on Theoretical and Physical Chemistry. Part 1. Chemical Dynamics* (Edward Arnold, London, 1898).
85. Arrhenius, S. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Z. Phys. Chem.* **4**, 226–248 (1889).
86. Tjoelker, M. G., Oleksyn, J. & Reich, P. B. Modelling respiration of vegetation: evidence for a temperature-dependent Q₁₀. *Glob. Change Biol.* **7**, 223–230 (2001).
87. Imriskova, I. *et al.* Biochemical characterization of the glucose kinase from *Streptomyces coelicolor* compared to *Streptomyces peucetius* var. *caesius*. *Res. Microbiol.* **156**, 361–366 (2005).
88. Michaelis, L. & Menten, M. L. Die kinetik der invertin wirkung. *Biochem. Z.* **49**, 334–336 (1913).
89. Atkin, O. K. & Tjoelker, M. G. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends Plant Sci.* **8**, 343–351 (2003).
90. Berry, J. A. & Raison, J. K. in *Physiological Plant Ecology I. Responses to the Physical Environment* (eds Lange, O. L. *et al.*) 277–388 (Springer, Berlin, 1981).
91. Davidson, E. A., Janssens, I. A. & Luo, Y. On the variability of respiration in terrestrial ecosystems: moving beyond Q₁₀. *Glob. Change Biol.* (in the press); published online 8 December 2005 (doi:10.1111/j.1365-2486.2005.01065.x).
92. Linkins, A. E., Melillo, J. M. & Sinsabaugh, R. L. in *Current Perspectives in Microbial Ecology* (eds Klug, M. J. & Reddy, C. A.) 572–579 (American Society for Microbiology, Washington DC, 1984).
93. Koch, K. E. Carbohydrate-modulated gene expression in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 509–540 (1996).
94. Sheen, J. Feedback control of gene expression. *Photosynth. Res.* **39**, 427–438 (1994).
95. Atkin, O. K., Bruhn, D., Hurry, V. M. & Tjoelker, M. G. The hot and the cold: unravelling the variable response of plant respiration to temperature. *Funct. Plant Biol.* **32**, 87–105 (2005).
96. Dalias, P., Anderson, J. M., Bottner, P. & Couëteux, M. M. Temperature response of carbon mineralization in conifer forest soils from different regional climates incubated under standard laboratory conditions. *Glob. Change Biol.* **7**, 181–192 (2001).
97. Dalias, P., Anderson, J. M., Bottner, P. & Couëteux, M. M. Long-term effects of temperature on carbon mineralisation processes. *Soil Biol. Biochem.* **33**, 1049–1057 (2001).
98. Hyvönen, R., Ågren, G. I. & Dalias, P. Analysing temperature response of decomposition of organic matter. *Glob. Change Biol.* **11**, 770–778 (2005).
99. Zak, D. R., Holmes, W. E., MacDonald, N. W. & Pregitzer, K. S. Soil temperature, matric potential, and the kinetics of microbial respiration and nitrogen mineralization. *Soil Sci. Soc. Am. J.* **63**, 575–584 (1999).
100. *Soil Organic Carbon Map* (US Department of Agriculture, Natural Resources Conservation Service, Soil Survey Division, World Soil Resources, Washington DC); (<http://soils.usda.gov/use/worldsoils/mapindex/soc.html>) (1997; revised 2000).

Acknowledgements We thank G. Ågren, R. Houghton, C. Potter, D. Sampson, J. Schimel, P. Smith and D. Thompson for comments on earlier drafts of this paper. We thank M. Ernst for assistance with graphics. E.A.D. acknowledges support from the Northeast Regional Center of the National Institute for Global Environmental Change of the Department of Energy and by the National Science Foundation. I.A.J. acknowledges the European Commission for support via CarboEurope-IP. Financial support does not constitute an endorsement by DOE or NSF of the views expressed in this article.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to E.A.D. (edavidson@whrc.org).

ARTICLES

Stochastic *spineless* expression creates the retinal mosaic for colour vision

Mathias F. Wernet^{1*†}, Esteban O. Mazzoni^{1*}, Arzu Çelik^{1*}, Dianne M. Duncan², Ian Duncan² & Claude Desplan¹

Drosophila colour vision is achieved by R7 and R8 photoreceptor cells present in every ommatidium. The fly retina contains two types of ommatidia, called 'pale' and 'yellow', defined by different rhodopsin pairs expressed in R7 and R8 cells. Similar to the human cone photoreceptors, these ommatidial subtypes are distributed stochastically in the retina. The choice between pale versus yellow ommatidia is made in R7 cells, which then impose their fate onto R8. Here we report that the *Drosophila* dioxin receptor *Spineless* is both necessary and sufficient for the formation of the ommatidial mosaic. A short burst of *spineless* expression at mid-pupation in a large subset of R7 cells precedes rhodopsin expression. In *spineless* mutants, all R7 and most R8 cells adopt the pale fate, whereas overexpression of *spineless* is sufficient to induce the yellow R7 fate. Therefore, this study suggests that the entire retinal mosaic required for colour vision is defined by the stochastic expression of a single transcription factor, *Spineless*.

The ability to discriminate between colours has evolved independently in vertebrates and invertebrates^{1,2}. However, despite the obvious differences in eye development and design, both flies and humans have developed retinal mosaics where classes of photoreceptor cells (PRs) with different spectral sensitivity are randomly distributed^{3,4}.

The compound eye of *Drosophila* consists of ~800 optical units (ommatidia), each containing eight PRs in addition to accessory cells⁵. In each ommatidium, the six 'outer PRs' (R1–R6) function like the vertebrate rod cells, as they are required for motion detection in dim light^{6,7}. These cells express the broad-spectrum rhodopsin, Rh1 (ref. 8). The 'inner PRs' (R7 and R8) may be viewed as the equivalent of the colour-sensitive vertebrate cone cells, which express a range of different rhodopsin molecules^{9–13}.

Ommatidial subset specification in *Drosophila*

The general rule of sensory receptor exclusion also applies to *Drosophila* ommatidia, where only one rhodopsin gene is expressed by a given PR¹⁴. The expression of inner PR rhodopsins can be used to distinguish three ommatidial subtypes^{15,16} (Supplementary Fig. 1a, b). Two of the subtypes are distributed randomly throughout the retina: ~30% of ommatidia express ultraviolet-sensitive Rh3 in R7 cells and blue-sensitive Rh5 in R8 cells, and therefore are specialized in the detection of short wavelengths ('pale' ommatidia, p; Fig. 1a, blue). The remaining ~70% express another ultraviolet-sensitive opsin (Rh4) in R7 and green-sensitive Rh6 in R8, making them more responsive to longer wavelengths ('yellow' ommatidia, y; Fig. 1a, yellow). The coupled expression of Rh3/Rh5 or Rh4/Rh6 within the same ommatidium results from communication between R7 and R8 (Supplementary Fig. 1b, c). In the dorsal rim area (DRA) (Fig. 1a, pink), a third type of ommatidia exists¹⁷ in which both R7 and R8 express ultraviolet-sensitive Rh3 (refs 18, 19). These ommatidia are used to detect the e-vector of polarized sunlight for orientation^{20,21}. Spatially localized polarized light detectors and stochastically distributed colour-sensitive ommatidia therefore

reflect two fundamentally different specification strategies that shape the retinal mosaic of *Drosophila*²².

The current model for specifying colour-sensitive ommatidia combines stochastic and instructive steps¹¹. First, a subset of R7 (pale R7, pR7) stochastically chooses Rh3 expression over the 'R7 default', Rh4. Second, these cells then impose the p fate (Rh5) onto R8 (pale R8, pR8) of the same ommatidium (Supplementary Fig. 1c).

In this study, we report the identification of *spineless* (*ss*) as a key regulatory gene for establishing the retinal mosaic. *ss* encodes the *Drosophila* homologue of the human arylhydrocarbon ('dioxin') receptor, a member of the bHLH-PAS (basic helix–loop–helix–Period–Arnt–Single-minded) family of transcription factors^{23,24}. At mid-pupation, *ss* is stochastically expressed in a majority of R7 that seem to correspond to the y subtype. *ss* is both necessary and sufficient to specify the yellow R7 (yR7) fate and subsequently the entire y ommatidia; pR7 cells are thus specified by default, and stochastic expression of *ss* represents the key regulatory event defining the retinal mosaic required for fly colour vision.

spineless is necessary for yellow ommatidia specification

We recently identified *homothorax* (*hth*) as the key regulatory gene necessary and sufficient for the specification of DRA ommatidia¹⁹. *ss* and *hth* cause similar homeotic phenotypes: that is, complete (*hth*) or partial (*ss*, '*aristapedia*') transformation of antennae into legs^{25,26}. Therefore, we tested for a potential role of *ss* in ommatidial subtype specification by generating whole-mutant eyes, as well as mitotic clones, lacking *ss* function using the null allele *ss*^{D115.7} and the *ey*-FLP/FRT technique^{27,28}. Owing to *ey*-FLP expression in the antennal imaginal disc, *ss* mutant flies showed a strong *aristapedia* phenotype (Supplementary Fig. 2a), but lacked any obvious morphological eye phenotype. However, expression of rhodopsin genes was severely affected. In wild-type eyes, Rh3 is found in ~30% of R7 cells, as well as in both R7 and R8 of DRA ommatidia (Fig. 1c, arrow), whereas the remaining ~70% of R7 contain Rh4 (Fig. 1c). In *ss* mutant eyes, Rh4 was completely absent, whereas Rh3 was expanded

¹Center for Developmental Genetics, Department of Biology, New York University, 100 Washington Place, New York, New York 10003, USA. ²Department of Biology, Washington University, Box 1229, 1 Brookings Drive, St Louis, Missouri 63130, USA. [†]Present address: Department of Neurobiology, Stanford University, Fairchild D200, 299 West Campus Drive, Stanford, California 94305, USA.

*These authors contributed equally to this work.

into all R7 cells (Fig. 1d). The total number of ommatidia was not reduced, indicating that R7 cells were mis-specified into pR7, rather than yR7 being specifically eliminated. *ss* mutant mitotic clones were morphologically wild type (Supplementary Fig. 2b); however, Rh3 was always present in mutant R7 cells (marked by the absence of β -galactosidase (β -gal) expression), whereas Rh4 was always lost (Fig. 1e).

To test whether the R7 *ss* phenotype was cell autonomous, we generated individual mutant R7 cells using the MARCM technique^{29,30}. All mutant R7 cells (marked by the presence of green fluorescent protein (GFP) expression) contained Rh3 and never Rh4, demonstrating that *ss* is required cell autonomously in R7 to induce Rh4 expression (Fig. 1f). DRA ommatidia were correctly specified in *ss* mutant eyes, as Rh3 was expressed normally in both DRA R7 and R8 cells (Fig. 1d, arrow). Therefore, *ss* is necessary for the establishment of the yR7 subtype without affecting PR fate specification (Fig. 1b and Supplementary Fig. 2c).

The ommatidial subtypes are first specified in R7, which then instruct R8 (ref. 16). Therefore, *ss* mutant eyes should exhibit a rhodopsin phenotype in R8. In wild types, ~30% of R8 cells contain Rh5, and the remaining ~70% contain Rh6 (Fig. 2c)^{11,12}. In *ss* mutant eyes, the large majority (up to 95%) of R8 contained Rh5 (Fig. 2a, d), with some R8 still containing Rh6. However, most of these remaining yR8 cells were located in the dorsal third of the eye (Fig. 2a). In this

part of the retina, instruction of pale R8 (pR8) by pR7 is less efficient, resulting in ommatidia with odd-coupled (Rh3/Rh6) rhodopsin expression (E.O.M. & C.D., manuscript in preparation). In *ss* mutants, the frequency of such ommatidia was significantly increased in the dorsal region (Fig. 2b).

To test whether the R8 opsin phenotype of *ss* mutants resulted from the inability of some mutant R7 cells to properly instruct R8, rather than from *ss* being directly required in R8, we generated *sevenless; spineless* (*sev; ss*) double-mutant eyes. These eyes, which lacked R7 cells, always exhibited the *sev* single-mutant phenotype (Fig. 2e), with virtually all R8 cells containing Rh6 (Fig. 2f). This indicates that *ss* is required in R7 for the formation of the yR7 subtype, and consequently for the formation of yR8, without being directly required in R8 PRs.

***spineless* induces the yellow ommatidial subtype**

We tested whether *ss* was also sufficient to induce the y ommatidial subtype (Fig. 3a). Overexpression of *ss* in all developing PRs using a strong *LGMR* (long *glass* multiple reporter)-Gal4 driver³¹ and *UAS-Ss* (*LGMR>ss* flies) resulted in a rough eye phenotype, as well as a dramatic rhodopsin phenotype: *Rh4* was activated in all PRs throughout the eye (R1–R6 as well as R7 and R8), as revealed by ectopic expression of an *Rh4*-GFP reporter in many PRs per ommatidium (Fig. 3c) compared with wild type (Fig. 3b). To avoid the

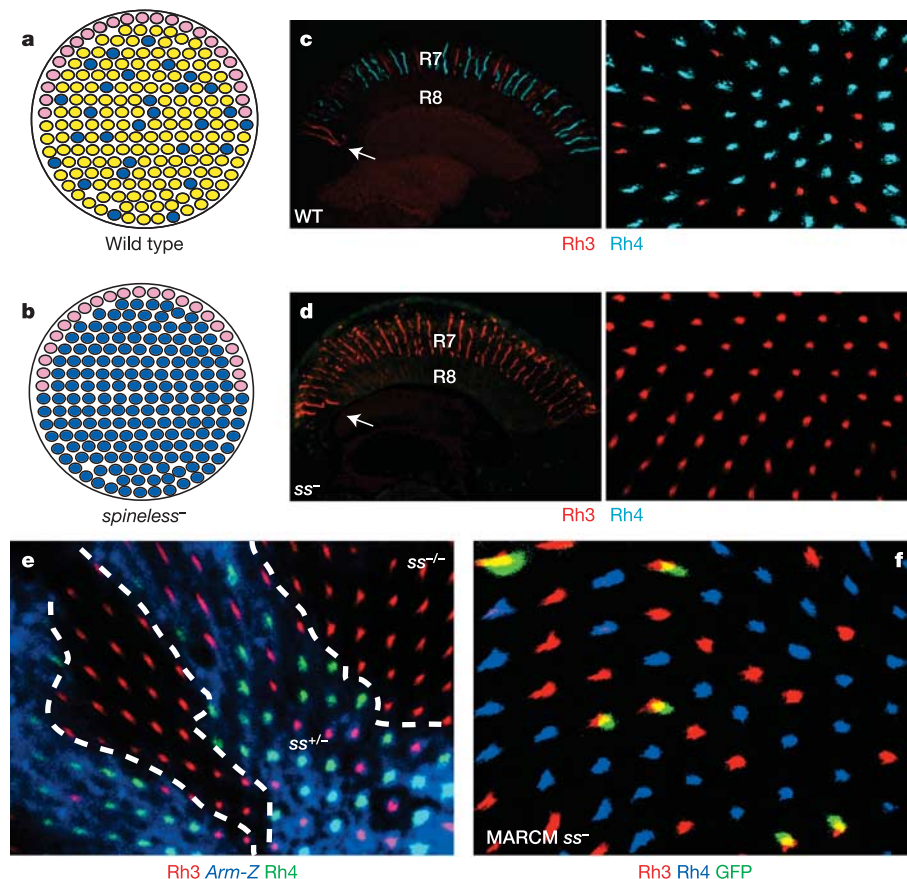


Figure 1 | The yR7 subtype is lost in *spineless* mutants. **a**, Three subtypes can be identified on the basis of molecular markers: 'pale' (blue), 'yellow' (yellow) and DRA (pink) ommatidia together form the wild-type retinal mosaic (schematic representation; dorsal is to the top). **b**, Schematic representation of the *ss* phenotype in R7 cells. **c**, Transverse section through a wild-type (WT) adult eye (left panel; dorsal is to the left). The arrow denotes the DRA. Ratio of R7 opsins in a wild-type whole-mount adult retina (right panel; dorsal is to the top) stained for Rh3 (red) and Rh4 (cyan). **d**, Transverse section through a *ss*^{D115.7} whole-mutant adult eye (left panel).

Rh3 (red) is expanded and Rh4 (cyan) is completely lost. Opsin expression in a mutant whole-mount adult retina is also shown (right panel). **e**, Whole-mount retina with mitotic clones lacking *Spineless* (marked by the absence of expression of the *armadillo-lacZ* (*Arm-Z*) construct, blue). Rh4 (green) is always absent from mutant clones. Rh3 expression is shown in red. **f**, Whole-mount retina with MARCM clones lacking *Spineless* (marked by the presence of GFP, green). Every GFP-positive cell expresses Rh3 (red), whereas Rh4 (blue) is always absent from mutant cells.

strong phenotype in the eye, we misexpressed *ss* using the weaker, variegated GMR driver, *sGMR* (short GMR)-Gal4³¹ (*sGMR>ss* flies). This led to strong ectopic induction of Rh4 in many PRs without severely affecting retinal morphology (Fig. 3d). This ectopic induction of Rh4 was also observed in *sev* mutants (Fig. 3f), and was thus independent of R7. Rh3 was still detected in some R7 in *sGMR>ss* flies, presumably due to the lack of variegated Gal4 expression in these cells (Fig. 3g, arrows), whereas Rh4 was expanded to some outer PRs. However, co-localization of Rh3 and Rh4 was never observed, confirming that gain of Rh4 in R7 cells always leads to the exclusion of Rh3. In contrast, gain of Rh4 in outer PRs did not lead to the exclusion of Rh1, as frequent coexpression of Rh1 and Rh4 was observed (Fig. 3i, arrows).

Using an *Rh4-lacZ* reporter construct in *LGMR>ss* flies, we found that β -gal-positive PR axons projected to both lamina and medulla, confirming the expansion of *Rh4* into outer PRs (Supplementary Fig. 3a). However, *Rh4*-expressing outer PRs were not transformed into genuine R7 cells, as they maintained their lamina projections (Supplementary Fig. 3a). Notably, DRA inner PRs were

the only cells not expressing *Rh4* (Supplementary Fig. 3a), suggesting that the DRA fate^{19,32}, specified by the gene *hth*, antagonizes *ss* function. Expression of *Rh3* and *Rh5* was completely lost (including in the DRA, where no rhodopsin was detected), while *Rh6* expression was found in most R8 cells (Supplementary Fig. 3b). This resulted in R8 coexpressing *Rh4* and *Rh6*, demonstrating that the 'one sensory receptor per cell' rule can be broken in *Drosophila* PRs, as has been shown in other insects^{14,33}. Therefore, ectopic induction of the γ R7 fate by *ss* specifically excludes the formation of pR7 cells. As a consequence, R8 cells expressing *Rh5* are not induced¹⁶, with most R8 expressing *Rh6*. *Rh6* was never found in outer PRs, supporting the hypothesis that *ss* is required only in R7 for the choice between *Rh3* and *Rh4*, and not directly in R8 for the *Rh6* choice (Supplementary Fig. 3b). In *LGMR>ss* flies, the specification of outer versus inner PRs (markers *spalt* and *seven up*; Supplementary Fig. 3c) or of R7 versus R8 (*prospero* and *senseless*; Supplementary Fig. 3d) was normal. Thus, *ss* acts by segregating ommatidial subtypes downstream of early PR specification events.

***spineless* can re-specify cell fate in R7 photoreceptor cells**

Colour PR cell fate determination seems to be a late event in PR differentiation. To test whether *ss* can transform the R7 fate at late stages of development, we used the *PanR7-Gal4* driver (which is also expressed in DRA R8 cells). Late mis-expression of *ss* induced the γ fate (Rh4) in all R7 cells, whereas Rh3 was absent (Fig. 4a, left panel). Opsin expression in the DRA was also altered, with Homothorax-positive cells (both R7 and R8) now expressing Rh4 (Supplementary Fig. 4). Hence, it is possible to reprogramme the R7 fate at later stages of differentiation, as *PanR7-Gal4* becomes activated at the time of rhodopsin expression. Surprisingly, expression of R8 rhodopsins outside the DRA was not affected, as the distribution of Rh5 and Rh6 resembled the wild type (Fig. 4a, right panel). As a result, many ommatidia manifested the very unusual coupling of Rh4 in R7 and Rh5 in R8 (Fig. 4b). Therefore, although *ss* is able to reprogramme all R7 late in development, R8 cannot revert their fate once they have been instructed to become pR8, and they maintain Rh5. We have recently identified two antagonistic genes expressed in either of the two R8 subtypes, which act together as a molecular consolidation system responsible for this inertia of R8 (ref. 34). To confirm that late expression of *ss* exclusively in R7 was sufficient to transform R7, *ss* was mis-expressed in *ss^{D115.7}* mutants using *PanR7-Gal4*. This was sufficient to induce Rh4 and to repress Rh3 (Fig. 4c, left panel). R8 were again not reprogrammed and exhibited the *ss* mutant phenotype, with many R8 cells expressing Rh5 (Fig. 4c, right panel).

Transient expression of *spineless* precedes R7 specification

All of the results presented above strongly indicated that *ss* must be expressed in the γ subtype of R7 at some point during pupal development. As several attempts to generate an anti-Spineless antibody had failed, we used *in situ* hybridization to detect *ss* messenger RNA in the retina at mid-pupation (Fig. 5a, b). At $\sim 50\%$ pupation, *ss* mRNA was detected in four neuronal cells per ommatidial cluster, one PR and three bristle cells (Fig. 5a, circles and asterisks). The PR was also labelled by anti-Prospero, confirming its identity as R7 (Fig. 5b). Although the expression levels of *ss* in bristle cells seemed uniformly high, levels of *ss* expression varied considerably among R7 cells, ranging from very faint to very strong in 60–80% of R7.

We also identified a 1.6 kilobase 'eye enhancer' fragment (*ss_{eye}*) within the *ss* promoter that drives PR-specific expression. After crossing *ss_{eye}-Gal4* to *UAS- β -gal::NLS* (nuclear localization sequence) reporters, PR-specific *ss* expression was first detected at mid-pupation—that is, approximately one day before rhodopsins are expressed, and before any visible molecular or morphological distinction between ommatidial subtypes (Fig. 5c). A single PR per ommatidium, which was identified as R7 through co-staining with Prospero, expressed *ss* (Supplementary Fig. 5a). Thus, the *ss_{eye}*

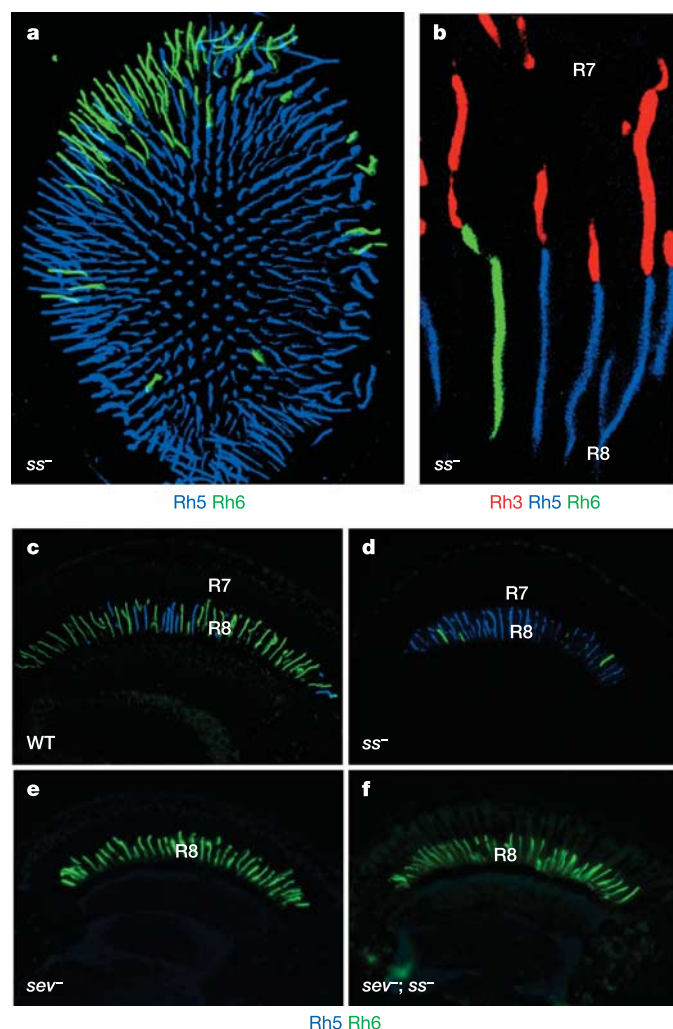


Figure 2 | R8 phenotype of *spineless* mutant eyes. **a**, Whole-mount retina from a *ss* mutant fly. The pR8 subtype (Rh5, blue) is expanded to almost all R8 cells (Rh6, green). **b**, Unusual mis-coupling of Rh3 (red) in R7 and Rh6 (green) in R8 of the same ommatidium is frequently observed in *ss* mutant retinas. Rh5 expression is shown in blue. **c–f**, Transverse sections stained for Rh5 (blue) and Rh6 (green): an adult wild-type eye (**c**), a *ss* mutant eye (**d**), a *sev* mutant eye (**e**), and a double-mutant (*sev; ss*) eye (**f**), which manifests the same R8 phenotype as *sev* mutants (expansion of γ R8 cells and loss of pR8 cells).

enhancer recapitulates endogenous *ss* expression in PRs. *ss* expression was detected in 60–80% of R7, correlating well with the distribution of Rh4 in adult retina. Like Rh4-expressing ommatidia, β -gal-positive ommatidia were more abundant in the dorsal half of the eye, and no β -gal expression was detected in the DRA (marked by Homothorax), where Rh4 is also never expressed (Supplementary Fig. 5b). *ss_{eye}*-Gal4 expression was detectable for only ~2 h at mid-pupation. Although it was not possible to directly co-stain for *ss* and Rh4 (which starts to be expressed one day later during pupation), it seems that at mid-pupation a short pulse of *ss* is deployed in a large subset of R7, which will become yR7.

spineless levels are crucial for specifying yellow ommatidia

We tested whether a short pulse of ectopic *ss* expression was able to modify the entire retinal mosaic, using a *heat shock*-Gal4 driver (*hs*-Gal4) to temporally control *ss* expression (*hs>ss* flies). A 30-min heat shock at ~50% pupation indeed resulted in an increase of Rh4 expression with a concomitant reduction of Rh3 in adults. The phenotype varied extensively, from only R7 cells expressing Rh4 (~25% of the flies analysed had Rh4 in most R7) (Fig. 6a), to almost every PR expressing Rh4 (Fig. 6b). In contrast, a 30-min pulse of *ss* in one-day-old adult flies had no effect (data not shown). Heat shocks during larval or early pupal stages were lethal. Thus, PRs are

extremely sensitive to a short pulse of *ss* during mid-pupation, at the time when endogenous *ss* is normally expressed.

To further study the mechanism of the stochastic choice between p and y ommatidia, we analysed the retinal mosaic in different mutant backgrounds. Flies heterozygous for *ss^{D115.7}* had fewer Rh4-expressing R7 cells (control: $66.1 \pm 3\%$; *ss* heterozygous: $53.8 \pm 3.5\%$; $P < 0.001$; values are mean $\pm$ s.d.). As the *ss^{D115.7}* allele only affects the *ss* coding sequence, heterozygous flies have two functional promoters, only one of which produces a functional protein, suggesting that the non-productive promoter might sequester limiting factor(s) that regulate(s) the expression levels of *ss*. If this hypothesis is correct, addition of extra copies of the *ss* promoter should have a similar effect. Indeed, the addition of two functional copies of the *ss* eye enhancer (*ss_{eye}*-Gal4) in an otherwise wild-type background also caused a significant reduction of the yR7 subtype ($50.2 \pm 2.4\%$, $P < 0.001$). Therefore, the level of Spineless expression is important for the induction of the yR7 fate, which is less efficient in cells where the amount of Spineless is reduced.

Conclusions

Retinal patterning in *Drosophila* reveals an original mechanism for how PR mosaics can be generated: stochastic expression of a single transcription factor (Spineless) acts as a binary switch that

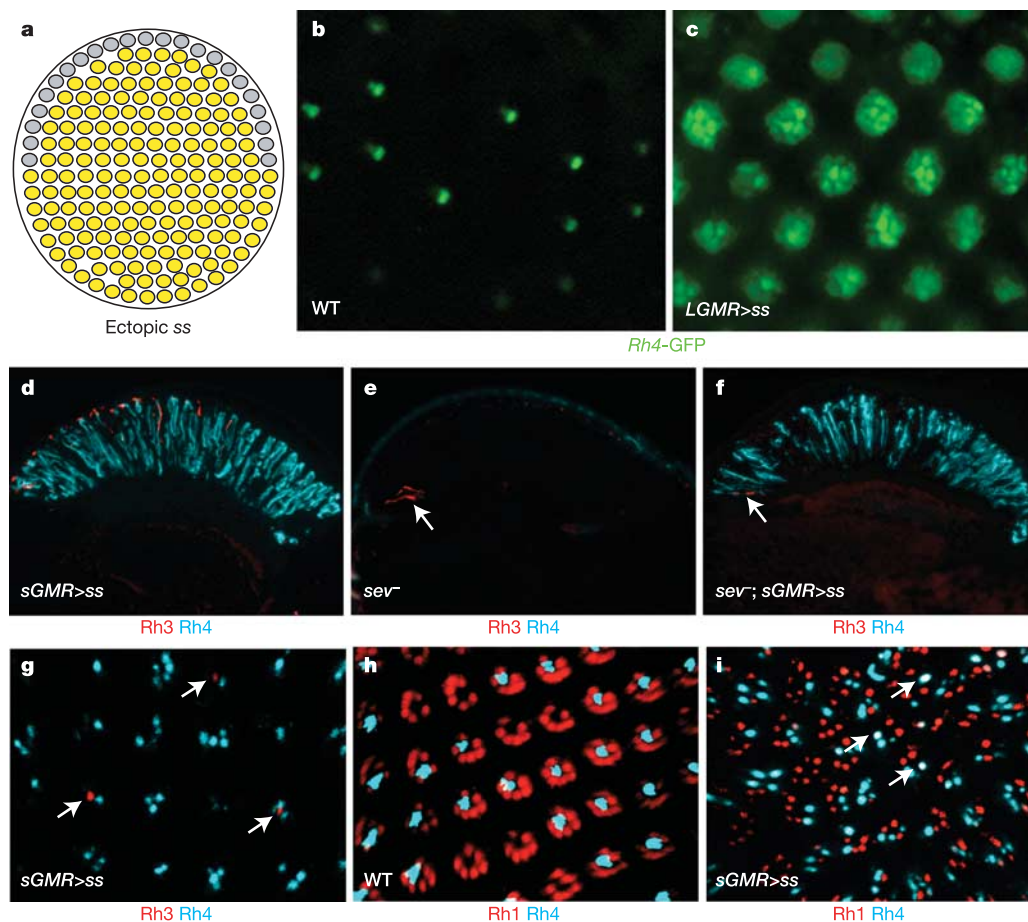


Figure 3 | *spineless* is sufficient to induce the yR7 fate. **a**, Summary of the *ss* gain-of-function phenotype. All R7 cells adopt the yR7 fate (yellow); the fate of DRA ommatidia is unclear (grey). **b**, Water immersion microscopy on living wild-type flies expressing the yR7-specific reporter *Rh4*-GFP. Expression is restricted to one inner PR in a large subset of ommatidia. **c**, *Rh4*-driven GFP expression is dramatically expanded in *LGMR>ss* flies, as visualized by water immersion. **d**, In *sGMR>ss* flies, *Rh4* (cyan) is expanded through the whole *ss*-overexpressing retina (compare with Fig. 1c). **e**, *Rh4* (cyan) is completely lost in *sev* mutants, although *Rh3* (red) is present in

DRA R8 cells (arrow). **f**, Overexpression of *ss* (*sGMR>ss*) leads to ectopic *Rh4* (cyan) in *sev* mutants. *Rh3* (red) is restricted to DRA R8 cells (arrow). **g**, In whole-mount retinas from *sGMR>ss* flies, variegated expression of *ss* leads to the expansion of *Rh4* (cyan) into a variable number of PRs per ommatidium; coexpression with *Rh3* (red) is never observed (arrows). **h**, In wild-type eyes, the outer-PR opsin *Rh1* (red) and *Rh4* (cyan) never co-localize in the same PR cell. **i**, Ectopic expression of *ss* in *sGMR>ss* flies leads to massive expansion of *Rh4* (cyan), with some co-localization (arrows) with the outer-PR opsin *Rh1* (red).

transforms the seemingly homogeneous compound eye into a mosaic, distinguishing p and y subtypes. However, subtype specification and rhodopsin expression can be separated, as *ss* expression in yR7 has ceased well before the time of rhodopsin expression (Fig. 6c). Additional factors are therefore required downstream of *ss* to ensure expression of adult p- and y-specific markers such as rhodopsins and additional screening pigments⁴. We propose a revised two-step model for the stochastic specification of p and y ommatidia (Fig. 6d). First, R7 are stochastically divided into two subtypes by the induction of *ss* in yR7. *ss*-positive R7 express *Rh4*, whereas the remaining R7 choose the pR7 fate and express *Rh3* by default (Fig. 6d, left). Second, only those R7 cells that did not express *ss* (pR7) retain the ability to induce the pR8 fate (*Rh5*), whereas yR8 express *Rh6* by default (Fig. 6d, right). The 'default states' of R7 (*Rh3*) and R8 (*Rh6*) therefore belong to opposite subtypes. Expression of R8 rhodopsin genes is maintained by a bistable regulatory loop containing the genes *warts* and *melted*³⁴. Notably, the localized specification of polarization-sensitive DRA ommatidia by *hth* antagonizes the stochastic choice executed by *ss*, placing these two genes into a new regulatory relationship during retinal patterning. Therefore, the role of the transcription factor Spineless is to generate the retinal mosaic required for fly colour vision by distinguishing yR7 from pR7 cell fates, and preventing R7 from instructing the underlying R8 cells.

Mosaic expression of sensory receptors has been described in detail

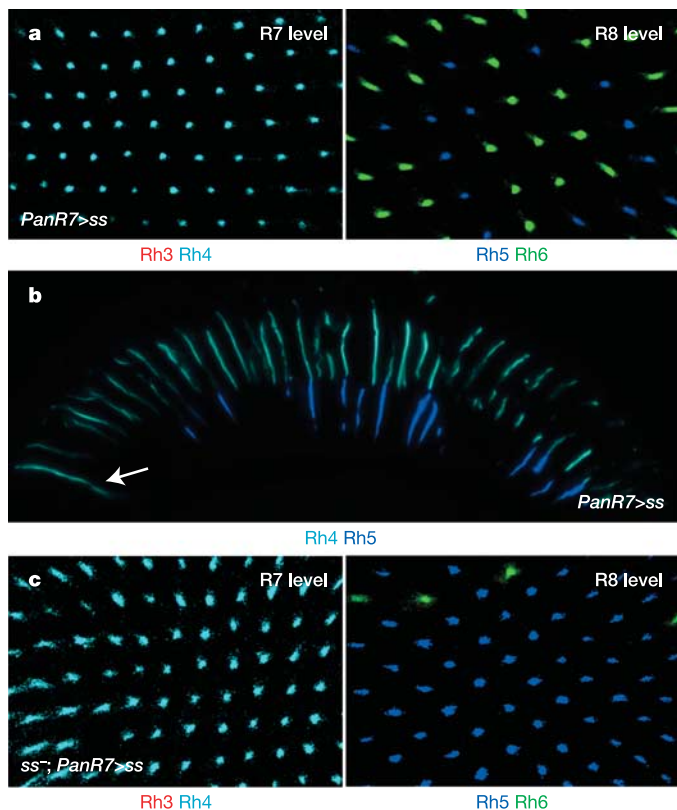


Figure 4 | Control of PR cell fates by *spineless*. **a**, In whole-mount retinas from *PanR7>ss* flies, all R7 cells express *Rh4* (cyan), whereas *Rh3* (red) is absent (left panel). Late expression of *ss* in R7 has no effect on the ratio between *Rh5* (blue) and *Rh6* (green) expression (right panel). **b**, Section of a *PanR7>ss* eye. As the fate of R8 is not affected by late *ss* expression in R7 cells, there is a great number of 'odd-coupled' *Rh4* (cyan)/*Rh5* (blue) ommatidia. The arrow indicates the DRA. **c**, In whole-mount retinas from *PanR7>ss* flies on a *ss* mutant background, *Rh3* (red) is not expressed, while every R7 cell expresses *Rh4* (cyan) (left panel). Late overexpression of *ss* in all R7 cells on a *ss* mutant background does not lead to a reprogramming of R8 cells as most R8 cells express *Rh5* (blue) and few dorsal R8 express *Rh6* (green) (right panel).

for the olfactory system of both vertebrates³⁵ and insects³⁶, and random PR mosaics have been described for humans³ and amphibians³⁷, as well as insects^{4,15,38–40}. Two transcription factors have been shown to regulate the specification of blue versus red/green cone cell fates in mammals. Upon mutation of either—the human nuclear receptor NR2E3 (also known as PNR) or the rodent thyroid hormone

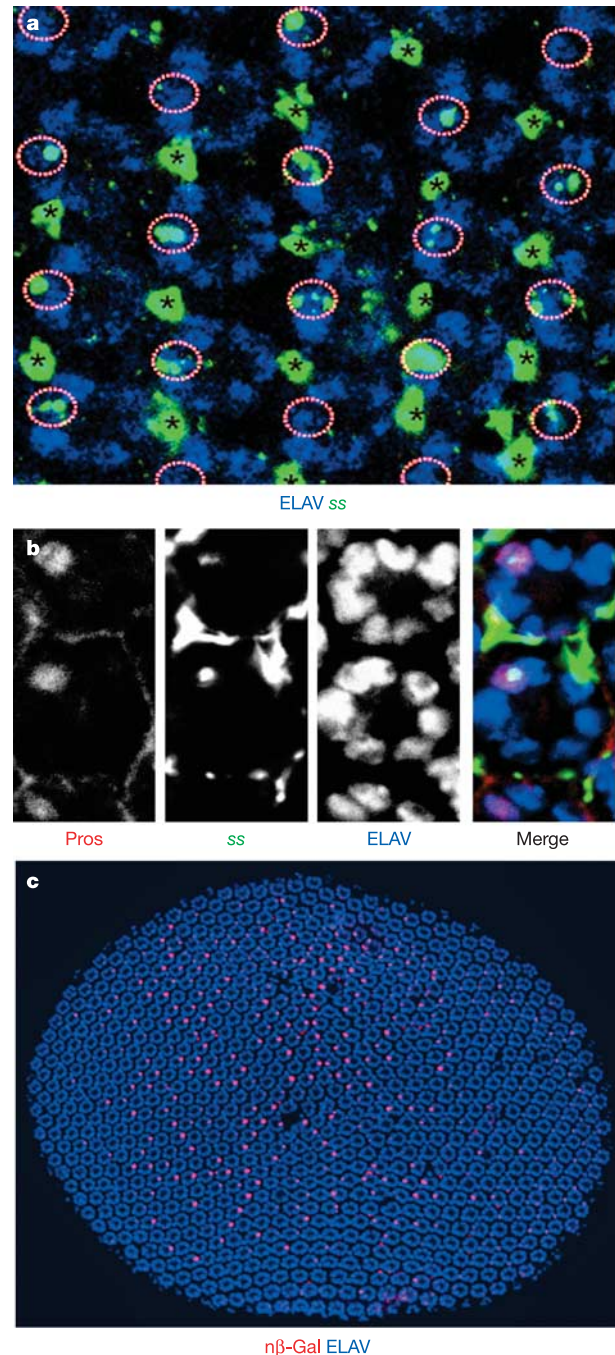


Figure 5 | Mosaic expression of *spineless* in the developing retina. **a**, *In situ* hybridization of a whole-mount pupal retina with an antisense *ss* probe (green) and an ELAV antibody (blue). *ss* expression can be observed in all bristle cells (asterisks), as well as in one PR (circles) per ommatidium in only 60–80% of all ommatidia. Note that *ss* levels vary from cell to cell. **b**, The PR cell positive for *ss* (green) is identified as an R7 cell by co-staining with the R7-specific marker Prospero (Pros, red). Neurons are marked with ELAV (blue). **c**, Nuclear β -gal ($n\beta$ -Gal, red) driven under the control of *ss*_{eye}-Gal4 reveals mosaic expression of *ss* in one cell per cluster. Neurons are marked with ELAV (blue). Dorsal is to the left in all panels.

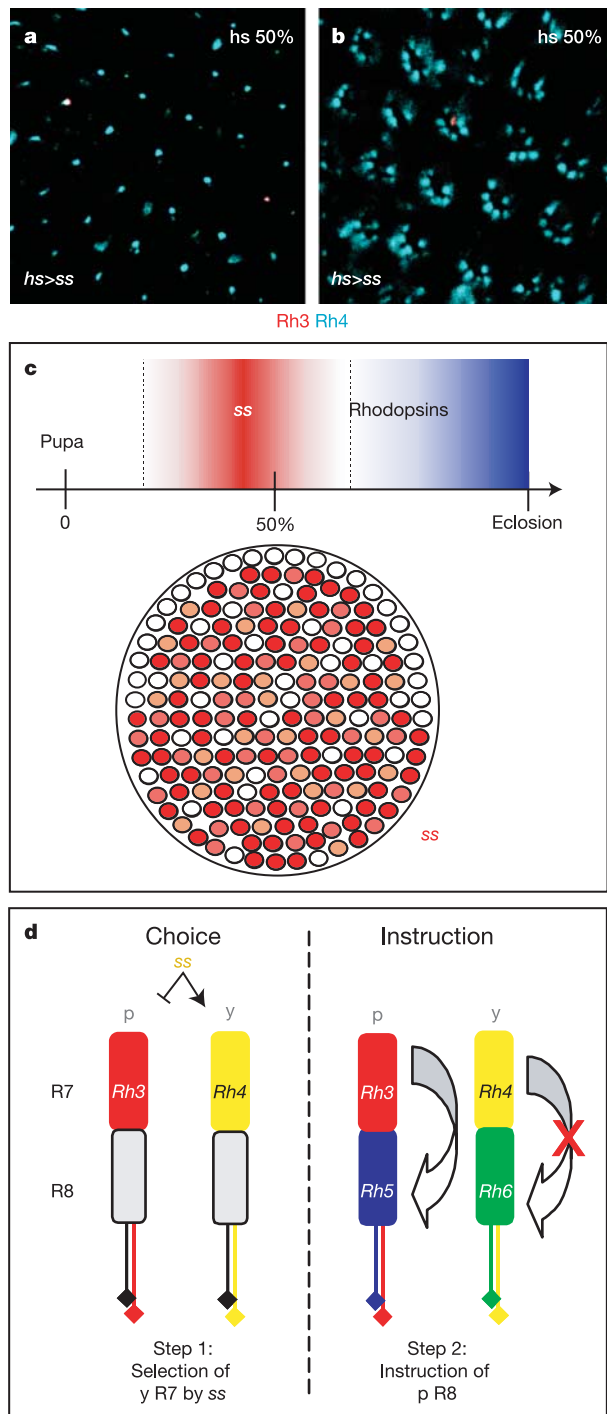


Figure 6 | Stochasticity of *ss* expression and a revised model for retinal patterning in *Drosophila*. **a, b,** The effect of a 30-min pulse of *ss* expression at ~50% pupation leads to an almost 100% transformation of R7 to the y fate (**a**), or almost every PR (**b**). **c,** Top: transient expression of *ss_{eye}*-Gal4 (red) during pupation before the onset of opsin expression (blue). Bottom: variable expression of *ss* (different tones of red) in R7 cells. **d,** Left: the *ss* data suggest that ~70% of the R7 cells get promoted into the yR7 fate (*Rh4*, yellow) by expressing *ss*. The pR7 subtype (*Rh3*, red) therefore represents the R7 'default state'. Right: in *ss*-positive yR7 cells, the ability to communicate with the underlying R8 is abolished, resulting in y ommatidia as the R8 default state is expression of *Rh6* (green). Only pR7 retain the competence to instruct the pR8 fate (*Rh5*, blue).

$\beta 2$ receptor—the number of blue cones is dramatically increased at the expense of green cones^{41,42}, leading to 'enhanced S-cone syndrome'. It should be noted that this retinal phenotype bears important similarity to the altered ommatidial mosaic in *Drosophila ss* mutants, where long wavelength-sensitive y ommatidia are lost at the expense of the short wavelength-sensitive p type.

The stochastic cell fate choice occurs at the level of the *ss* promoter: the very short pulse of *ss* expression at mid-pupation is not only controlled temporally, but its levels are also critical, and only ~70% of R7 receive enough Spineless to commit to the yR7 fate. Elucidating the mechanism that controls *ss* expression will shed some light into the fascinating process of stochastic gene expression, and the identification of its downstream targets will provide insights into consolidation and maintenance of cell fates.

METHODS

Drosophila strains and crosses, constructs, generation of transgenic flies by germ line transformation, antibody staining on mid-pupal and adult whole-mounted or cryo-sectioned retinas, *in situ* hybridization on pupal retinas, MARCM, cornea neutralization, and adult eye plastic sections were all performed by standard methods, and are described in detail in Supplementary Information.

Received 18 September 2005; accepted 30 January 2006.

- Gegenfurtner, K. R. & Kiper, D. C. Color vision. *Annu. Rev. Neurosci.* **26**, 181–206 (2003).
- Briscoe, A. D. & Chittka, L. The evolution of color vision in insects. *Annu. Rev. Entomol.* **46**, 471–510 (2001).
- Roorda, A. & Williams, D. R. The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–522 (1999).
- Franceschini, N., Kirschfeld, K. & Minke, B. Fluorescence of photoreceptor cells observed *in vivo*. *Science* **213**, 1264–1267 (1981).
- Wolff, T. & Ready, D. F. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 1277–1325 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993).
- Heisenberg, M. & Buchner, E. The role of retinula cell types in visual behavior of *Drosophila melanogaster*. *J. Comp. Physiol.* **117**, 127–162 (1977).
- Miller, G., Hansen, K. & Stark, W. Phototaxis in *Drosophila*: R1–R6 input and interaction among ocellar and compound eye receptors. *J. Insect Physiol.* **27**, 813–819 (1981).
- Zuker, C. S., Cowman, A. F. & Rubin, G. M. Isolation and structure of a rhodopsin gene from *D. melanogaster*. *Cell* **40**, 851–858 (1985).
- Montell, C., Jones, K., Zuker, C. & Rubin, G. A second opsin gene expressed in the ultraviolet-sensitive R7 photoreceptor cells of *Drosophila melanogaster*. *J. Neurosci.* **7**, 1558–1566 (1987).
- Fryxell, K. J. & Meyerowitz, E. M. An opsin gene that is expressed only in the R7 photoreceptor cell of *Drosophila*. *EMBO J.* **6**, 443–451 (1987).
- Chou, W. H. *et al.* Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17**, 1101–1115 (1996).
- Papatsenko, D., Sheng, G. & Desplan, C. A new rhodopsin in R8 photoreceptors of *Drosophila*: evidence for coordinate expression with Rh3 in R7 cells. *Development* **124**, 1665–1673 (1997).
- Huber, A. *et al.* Molecular cloning of *Drosophila* Rh6 rhodopsin: the visual pigment of a subset of R8 photoreceptor cells. *FEBS Lett.* **406**, 6–10 (1997).
- Mazzoni, E. O., Desplan, C. & Celik, A. 'One receptor' rules in sensory neurons. *Dev. Neurosci.* **26**, 388–395 (2004).
- Feiler, R. *et al.* Ectopic expression of ultraviolet-rhodopsins in the blue photoreceptor cells of *Drosophila*: visual physiology and photochemistry of transgenic animals. *J. Neurosci.* **12**, 3862–3868 (1992).
- Chou, W. H. *et al.* Patterning of the R7 and R8 photoreceptor cells of *Drosophila*: evidence for induced and default cell-fate specification. *Development* **126**, 607–616 (1999).
- Wada, S. Spezielle randzonale ommatidien der fliegen (Diptera: Brachycera): Architektur und verteilung in den komplexaugen. *Z. Morphol. Tiere* **77**, 87–125 (1974).
- Fortini, M. E. & Rubin, G. M. Analysis of *cis*-acting requirements of the *Rh3* and *Rh4* genes reveals a bipartite organization to rhodopsin promoters in *Drosophila melanogaster*. *Genes Dev.* **4**, 444–463 (1990).
- Wernet, M. F. *et al.* Homothorax switches function of *Drosophila* photoreceptors from color to polarized light sensors. *Cell* **115**, 267–279 (2003).
- von Philipsborn, A. & Labhart, T. A behavioral study of polarization vision in the fly, *Musca domestica*. *J. Comp. Physiol.* **167**, 737–743 (1990).
- Wolf, R., Gebhardt, B., Gademann, R. & Heisenberg, M. Polarization sensitivity of course control in *Drosophila melanogaster*. *J. Comp. Physiol.* **139**, 177–191 (1980).
- Wernet, M. F. & Desplan, C. Building a retinal mosaic: cell-fate decision in the fly eye. *Trends Cell Biol.* **14**, 576–584 (2004).

23. Emmons, R. B. *et al.* The Spineless-Aristapedia and Tango bHLH-PAS proteins interact to control antennal and tarsal development in *Drosophila*. *Development* **126**, 3937–3945 (1999).
24. Duncan, D. M., Burgess, E. A. & Duncan, I. Control of distal antennal identity and tarsal development in *Drosophila* by *spineless-aristapedia*, a homolog of the mammalian dioxin receptor. *Genes Dev.* **12**, 1290–1303 (1998).
25. Burgess, E. A. & Duncan, I. Direct control of antennal identity by the *spineless-aristapedia* gene of *Drosophila*. *Mol. Gen. Genet.* **221**, 347–357 (1990).
26. Casares, F. & Mann, R. S. Control of antennal versus leg development in *Drosophila*. *Nature* **392**, 723–726 (1998).
27. Stowers, R. S. & Schwarz, T. L. A genetic method for generating *Drosophila* eyes composed exclusively of mitotic clones of a single genotype. *Genetics* **152**, 1631–1639 (1999).
28. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
29. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
30. Pignoni, F. *et al.* The eye-specification proteins So and Eya form a complex and regulate multiple steps in *Drosophila* eye development. *Cell* **91**, 881–891 (1997).
31. Moses, K. & Rubin, G. M. Glass encodes a site-specific DNA-binding protein that is regulated in response to positional signals in the developing *Drosophila* eye. *Genes Dev.* **5**, 583–593 (1991).
32. Tomlinson, A. Patterning the peripheral retina of the fly: decoding a gradient. *Dev. Cell* **5**, 799–809 (2003).
33. Tahayato, A. *et al.* Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the *Drosophila* retina. *Dev. Cell* **5**, 391–402 (2003).
34. Mikeladze-Dvali, T. *et al.* The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors. *Cell* **122**, 775–787 (2005).
35. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
36. Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725–736 (1999).
37. Sherry, D. M., Bui, D. D. & Degrip, W. J. Identification and distribution of photoreceptor subtypes in the neotenic tiger salamander retina. *Vis. Neurosci.* **15**, 1175–1187 (1998).
38. White, R. H., Xu, H., Munch, T. A., Bennett, R. R. & Grable, E. A. The retina of *Manduca sexta*: rhodopsin expression, the mosaic of green-, blue- and UV-sensitive photoreceptors, and regional specialization. *J. Exp. Biol.* **206**, 3337–3348 (2003).
39. Schmitt, A., Vogt, A., Friedmann, K., Paulsen, R. & Huber, A. Rhodopsin patterning in central photoreceptor cells of the blowfly *Calliphora vicina*: cloning and characterization of *Calliphora* rhodopsins Rh3, Rh5 and Rh6. *J. Exp. Biol.* **208**, 1247–1256 (2005).
40. Wakakuwa, M., Kurasawa, M., Giurfa, M. & Arikawa, K. Spectral heterogeneity of honeybee ommatidia. *Naturwissenschaften* **92**, 464–467 (2005).
41. Haider, N. B. *et al.* Mutation of a nuclear receptor gene, *NR2E3*, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nature Genet.* **24**, 127–131 (2000).
42. Ng, L. *et al.* A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nature Genet.* **27**, 94–98 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Britt, R. Carthew, S. Crews, R. Mann, G. Struhl, J. Treisman, the Hybridoma Bank, and the Bloomington Stock Center for flies and reagents. We also would like to thank J. Treisman and J. Blau for comments on the manuscript. We thank N. Franceschini, as well as all past and present members of the Desplan laboratory, for suggestions throughout the project and their comments on the manuscript. M.F.W. was supported by the Boehringer Ingelheim Fonds and the Studienstiftung des Deutschen Volkes, E.O.M. was supported by the Dean's Dissertation Fellowship from NYU. This work was funded by a grant from the National Eye Institute/NIH to C.D. and conducted in a facility constructed with the support of a Research Facilities Improvement Grant from the NCRR, NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D. (claudeddesplan@nyu.edu).

A photometric redshift of $z = 6.39 \pm 0.12$ for GRB 050904

J. B. Haislip¹, M. C. Nysewander¹, D. E. Reichart¹, A. Levan², N. Tanvir², S. B. Cenko³, D. B. Fox⁴, P. A. Price⁵, A. J. Castro-Tirado⁶, J. Gorosabel⁶, C. R. Evans¹, E. Figueredo^{7,8}, C. L. MacLeod¹, J. R. Kirschbrown¹, M. Jelinek⁶, S. Guziy⁶, A. de Ugarte Postigo⁶, E. S. Cypriano^{8,9}, A. LaCluyze¹, J. Graham¹⁰, R. Priddey², R. Chapman², J. Rhoads¹¹, A. S. Fruchter¹¹, D. Q. Lamb¹², C. Kouveliotou¹³, R. A. M. J. Wijers¹⁴, M. B. Bayliss^{1,12}, B. P. Schmidt¹⁵, A. M. Soderberg³, S. R. Kulkarni³, F. A. Harrison¹⁶, D. S. Moon³, A. Gal-Yam³, M. M. Kasliwal³, R. Hudec¹⁷, S. Vitek¹⁸, P. Kubanek¹⁹, J. A. Crain¹, A. C. Foster¹, J. C. Clemens¹, J. W. Bartelme¹, R. Canterna²⁰, D. H. Hartmann²¹, A. A. Henden²², S. Klose²³, H.-S. Park²⁴, G. G. Williams²⁵, E. Rol²⁶, P. O'Brien²⁶, D. Bersier²⁷, F. Prada⁶, S. Pizarro⁸, D. Maturana⁸, P. Ugarte⁸, A. Alvarez⁸, A. J. M. Fernandez⁶, M. J. Jarvis²⁸, M. Moles⁶, E. Alfaro⁶, K. M. Ivarsen¹, N. D. Kumar¹, C. E. Mack¹, C. M. Zdarowicz¹, N. Gehrels²⁹, S. Barthelmy²⁹ & D. N. Burrows⁴

Gamma-ray bursts (GRBs) and their afterglows are the most brilliant transient events in the Universe. Both the bursts themselves and their afterglows have been predicted to be visible out to redshifts of $z \approx 20$, and therefore to be powerful probes of the early Universe^{1,2}. The burst GRB 000131, at $z = 4.50$, was hitherto the most distant such event identified³. Here we report the discovery of the bright near-infrared afterglow of GRB 050904 (ref. 4). From our measurements of the near-infrared afterglow, and our failure to detect the optical afterglow, we determine the photometric redshift of the burst to be $z = 6.39^{+0.11}_{-0.12}$ (refs 5–7). Subsequently, it was measured⁸ spectroscopically to be

$z = 6.29 \pm 0.01$, in agreement with our photometric estimate. These results demonstrate that GRBs can be used to trace the star formation, metallicity, and reionization histories of the early Universe.

At 01:51:44 Universal Time (UT) on 4 September 2005, Swift's Burst Alert Telescope (BAT) detected GRB 050904 and 81 seconds later a 4'-radius localization was distributed to observers on the ground. Swift's X-Ray Telescope (XRT) automatically slewed to the BAT localization and 76 minutes after the burst a 6"-radius XRT localization was distributed.⁹

Over the next few hours, we observed the XRT localization at both



Figure 1 | NIR and visible-light images of the field of GRB 050904. Left, NIR discovery image of the bright ($J = 17.36 \pm 0.04$ mag) afterglow of GRB 050904 from 4.1-m SOAR on top of Cerro Pachon, Chile. Middle, near-simultaneous non-detection of the afterglow at visible wavelengths (unfiltered, calibrated to $R_c > 20.1$ mag) from one of the six 0.41-m PROMPT telescopes that we are building on top of Cerro Tololo, which is only 10 km away from Cerro Pachon. Right, colour composite ($r'i'z'$) image of the afterglow 3.2 days after the burst from 8.1-m Gemini South, which is also on top of Cerro Pachon.

¹Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Campus Box 3255, Chapel Hill, North Carolina 27599, USA. ²Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK. ³Department of Astronomy, California Institute of Technology, Pasadena, California 91125, USA. ⁴Department of Astronomy and Astrophysics, 525 Davey Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁵Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ⁶Instituto de Astrofísica de Andalucía, PO Box 3.004, 18.080 Granada, Spain. ⁷Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matao 1226, 05508-900 São Paulo, SP, Brazil. ⁸Southern Observatory for Astrophysical Research, Casilla 603, La Serena, Chile. ⁹Laboratorio Nacional de Astrofísica, CP 21, 37500-000 Itajuba, MG, Brazil. ¹⁰Department of Astronomy, 601 Campbell Hall, University of California, Berkeley, California 94720, USA. ¹¹Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ¹²Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA. ¹³NASA Marshall Space Flight Center, National Space Science Technology Center, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ¹⁴Astronomical Institute "Anton Pannekoek", University of Amsterdam and Center for High-Energy Astrophysics, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands. ¹⁵Mount Stromlo and Siding Spring Observatories, Private Bag, Weston Creek PO, Canberra, ACT 2611, Australia. ¹⁶Space Radiation Laboratory, California Institute of Technology, MC 220-47, Pasadena, California 91125, USA. ¹⁷Astronomical Institute, Academy of Sciences of the Czech Republic, 25165 Ondřejov, Czech Republic. ¹⁸Faculty of Electrotechnics, Czech Technical University, 121 35 Praha, Czech Republic. ¹⁹Integral Science Data Center, Chemin d'Ecogia 16, CH-1290 Versoix, Switzerland. ²⁰Department of Physics and Astronomy, University of Wyoming, PO Box 3905, Laramie, Wyoming 82072, USA. ²¹Clemson University, Department of Physics and Astronomy, Clemson, South Carolina 29634, USA. ²²American Association of Variable Star Observers, 25 Birch Street, Cambridge, Massachusetts 02138, USA. ²³Thüringer Landessternwarte Tautenburg, Sternwarte 5, D-07778 Tautenburg, Germany. ²⁴Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA. ²⁵Multiple Mirror Telescope Observatory, University of Arizona, Tucson, Arizona 85721, USA. ²⁶Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ²⁷Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, UK. ²⁸Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK. ²⁹NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

Table 1 | Observations of the afterglow of GRB 050904

Date (UT)	Mean Δt	Filter	Zero point (Jy)	Magnitude*	Telescope
Sep 4.0795	2.80 min	R	3,105	>18.2	0.30-m BOOTES-1B
Sep 4.0821	6.46 min	R	3,105	>18.3	0.30-m BOOTES-1B
Sep 4.0868	13.22 min	R	3,105	>19.2	0.30-m BOOTES-1B
Sep 4.0956	25.95 min	R	3,105	>19.5	0.30-m BOOTES-1B
Sep 4.1151	53.96 min	R	3,105	>19.9	0.30-m BOOTES-1B
Sep 4.1535	109.30 min	R	3,105	>21.0	3.5-m Calar Alto
Sep 4.206	3.07 h	J	1,614	17.36 ± 0.04	4.1-m SOAR
Sep 4.213	3.25 h	J	1,614	17.35 ± 0.04	4.1-m SOAR
Sep 4.220	3.42 h	J	1,614	17.61 ± 0.04	4.1-m SOAR
Sep 4.248	4.08 h	Z	3,631	>18.8	60-inch Palomar
Sep 4.355	6.66 h	R	3,105	>22.3	60-inch Palomar
Sep 4.366	6.91 h	Unfiltered, calibrated to R_c	3,105	>20.1	0.41-m PROMPT-5
Sep 4.390	7.49 h		1,614	18.66 ± 0.15	4.1-m SOAR
Sep 4.402	7.78 h		676	16.77 ± 0.07	4.1-m SOAR
Sep 4.416	8.12 h		3,631	>21.1	60-inch Palomar
Sep 4.486	9.79 h	H	1,049	18.17 ± 0.06	3.8-m UKIRT
Sep 4.488	9.86 h	J	1,614	19.02 ± 0.06	3.8-m UKIRT
Sep 4.502	10.18 h	K	676	17.38 ± 0.06	3.8-m UKIRT
Sep 4.518	10.57 h	K'	676	17.55 ± 0.03	3.0-m IRTF
Sep 4.551	11.35 h	Z	2,270	22.08 ± 0.16	3.8-m UKIRT
Sep 4.565	11.69 h	J	1,614	19.25 ± 0.07	3.8-m UKIRT
Sep 5.198	26.90 h	Y	2,060	20.42 ± 0.26	4.1-m SOAR
Sep 5.246	28.03 h	J	1,614	20.16 ± 0.17	4.1-m SOAR
Sep 5.322	29.87 h	I_c	2,433	>20.2	0.41-m PROMPT-3 + 0.41-m PROMPT-5
Sep 6.30	2.22 day	J	1,614	20.60 ± 0.23	4.1-m SOAR
Sep 6.35	2.27 day	Y	2,060	20.98 ± 0.34	4.1-m SOAR
Sep 7.21	3.13 day	i'	3,631	>25.4	8.1-m Gemini South
Sep 7.23	3.15 day	r'	3,631	>26.5	8.1-m Gemini South
Sep 7.24	3.16 day	z'	3,631	23.36 ± 0.14	8.1-m Gemini South

* Error bars are 1σ and upper limits are 3σ .

We calibrated the $r'i'z'$ measurements using stellar Sloan Digital Sky Survey (SDSS) sources and derived R_cI_c field calibrations from the SDSS field calibrations²⁴. We obtained YJHK_sK field calibrations using SOAR and ZJHK field calibrations using UKIRT. The JHK field calibrations are in agreement with each other and with the 2-Micron All-Sky Survey (2MASS). However, the UKIRT Z-band measurement, which we obtained 11.4 h after the burst, is a factor of three below the fitted model (Fig. 2). The UKIRT WFCAM Z bandpass was designed to match the effective wavelength of the SDSS z' bandpass (0.876 versus 0.887 μm), but with a rectangular profile. The standard deviation of the magnitude differences for all stellar SDSS sources in the UKIRT Z field and the Gemini South z' field, which we obtained 3.2 days after the burst, is only 0.064 mag. When converting from magnitudes to spectral fluxes, we used the correct zero points for Z and z' , respectively. When fitting to these spectral fluxes, we used the actual UKIRT WFCAM Z and Gemini South GMOS-S z' bandpasses. No modification of the model spectrum (for example, dust extinction²¹, molecular hydrogen absorption²⁵, or the Ly α damping wing) appears to be able to accommodate both measurements simultaneously. Consequently, we conclude that this factor-of-three deficit is not only real but probably temporal in nature.

near-infrared (NIR) and visible wavelengths (Table 1). In the NIR, we discovered a bright ($J \approx 17.4$ mag at 3.1 hours after the burst) and fading source within the XRT localization using the 4.1-m Southern Observatory for Astrophysical Research (SOAR) telescope on top of Cerro Pachon in Chile (Fig. 1).⁴

However, at visible wavelengths we did not detect the afterglow to relatively deep limiting magnitudes using one of the six 0.41-m Panchromatic Robotic Optical Monitoring and Polarimetry Telescopes (PROMPT)¹⁰ that we are building on top of Cerro Tololo, which is only 10 km away from Cerro Pachon⁴, the 60-inch telescope at Palomar Observatory in California¹¹, and the 3.5-m telescope at Calar Alto Observatory in Spain. Nor did we detect the afterglow with the 0.30-m Burst Observer and Optical Transient Exploring System (BOOTES)¹² 1B telescope in El Arenosillo, Spain, which began imaging the field only 2.1 minutes after the burst¹³. This implies that the GRB either occurred at a very high redshift or was very heavily extinguished by dust⁴.

Between about 3 hours and about 0.5 days after the burst, the fading of the afterglow appears to be well described by a power law of index $-1.36^{+0.07}_{-0.06}$ (Fig. 2)^{5,6}. However, after this time the fading appears to have slowed to a temporal index of $-0.82^{+0.21}_{-0.08}$ (refs 7, 14, 15). A single power-law description is ruled out at the 3.7σ credible level. One possible explanation is that our initial SOAR observations caught the tail end of a reverse shock that had been stretched in time by a factor of 7.29 owing to cosmological time dilation. Even so, the reverse shock would still be at least a few times longer-lived in the source frame than the reverse shocks of GRBs 990123 and 021211 (refs 16, 17). Another possibility is that we are undersampling a light curve that is undergoing temporal variations, such as in the afterglows of GRBs 021004 and 030329 (refs 18, 19). Indeed, the X-ray afterglow is extremely variable at these times²⁰.

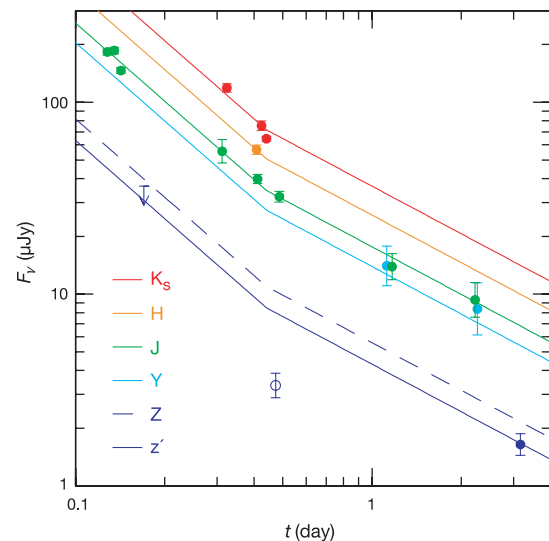


Figure 2 | NIR and z' -band light curves of the afterglow of GRB 050904 and our best-fit model. A single power-law description is ruled out at the 3.7σ credible level. Following the formalism of Frail *et al.*²⁶, given GRB 050904's redshift and fluence²⁷ the non-detection of a jet break in the light curve prior to 2.3 days after the burst implies that the opening/viewing angle of the jet is $\geq 3^\circ$ and that the total energy that was released in γ rays is $\geq 5 \times 10^{50}$ erg. The Z-band measurement (unfilled circle) is a factor of three below the fitted model, but this appears to be real and temporal in nature (Table 1). Error bars are s.e.m. Downward arrow indicates upper limit.

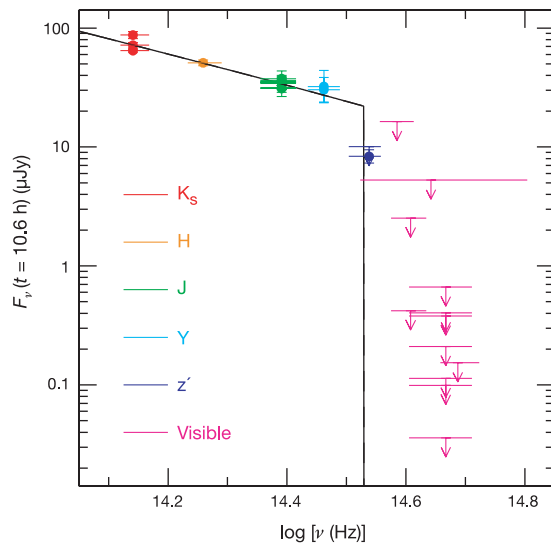


Figure 3 | Spectral flux distribution of the afterglow of GRB 050904 and our best-fit model. Measurements have been scaled to 10.6 hours after the burst using our best-fit light curve. We model the spectrum as a power law with negligible emission blueward of $\text{Ly}\alpha$. Shallower intrinsic power-law spectra can be inferred with the addition of source-frame dust. If one assumes that the jet is propagating through either a constant-density or wind-swept medium with the synchrotron electron cooling frequency either redward or blueward of the observed frequencies, the fitted temporal index ($-0.82^{+0.21}_{-0.08}$) implies an electron energy distribution index between $1.43^{+0.28}_{-0.11}$ and $2.09^{+0.28}_{-0.11}$ and an intrinsic spectral index between $-0.88^{+0.14}_{-0.05}$ and $-0.21^{+0.14}_{-0.05}$ (refs 28, 29). This is shallower than the fitted spectral index ($-1.25^{+0.15}_{-0.14}$), which suggests that source-frame dust is probably present. However, only a small amount is required to explain such a difference at these source-frame frequencies. This cannot explain the sharp drop in spectral flux in and blueward of the z' band^{5,21}. We take Galactic $E(B - V) = 0.060$ mag (ref. 30). Error bars are s.e.m. Downward arrows indicate upper limits.

Using these temporal indices to scale all of our measurements to a common time, except for a Z-band ($0.84\text{--}0.93\ \mu\text{m}$) measurement from 11.4 hours after the burst (Table 1), we plot the spectral flux distribution of the afterglow in Fig. 3. In the NIR, the spectral index is $-1.25^{+0.15}_{-0.14}$. However, the spectral index between NIR and visible wavelengths is steeper than -5.9 . This is too sharp a transition to be explained by dust extinction alone^{5,21}. However, a small amount of extinction cannot be ruled out and is probably present (Fig. 3).

Assuming negligible emission blueward of $\text{Ly}\alpha$, we measure a photometric redshift of $z = 6.39^{+0.11}_{-0.12}$ (refs 5–7), which is consistent with the spectroscopic redshift of $z = 6.29 \pm 0.01$ (ref. 8). For $H_0 = 71\ \text{km s}^{-1}\text{Mpc}^{-1}$, $\Omega_m = 0.27$, and $\Omega_\Lambda = 0.73$ (ref. 22), this corresponds to about 900 million years after the Big Bang, when the Universe was about 6% of its current age. The next-most-distant GRB that has been identified occurred at $z = 4.50$ (ref. 3), which was about 500 million years later, when the Universe was about 10% of its current age.

One of the most exciting aspects of this discovery is the brightness of the afterglow: extrapolating back to a few minutes after the burst, the afterglow must have been exceptionally bright redward of $\text{Ly}\alpha$ for the robotic 0.25-m TAROT telescope to detect it in unfiltered visible-light observations²³. Extrapolating our J-band light curve back to these times yields $J \approx 11\text{--}12$ mag. This suggests that by pairing visible-light robotic telescopes with NIR robotic telescopes, and these with larger telescopes that are capable of quick-response NIR spectroscopy, all preferably at the same site so that they are subject to the same observing conditions, at least some very-high-redshift afterglows will be discovered, identified, and their NIR spectrum

taken while they are still sufficiently bright to serve as a powerful probe of the conditions of the early Universe¹⁰.

Received 23 September; accepted 12 December 2005.

- Lamb, D. Q. & Reichart, D. E. Gamma-ray bursts as a probe of the very high redshift Universe. *Astrophys. J.* **536**, 1–18 (2000).
- Ciardi, B. & Loeb, A. Expected number and flux distribution of gamma-ray burst afterglows with high redshifts. *Astrophys. J.* **540**, 687–696 (2000).
- Andersen, M. I. *et al.* VLT identification of the optical afterglow of the gamma-ray burst GRB 000131 at $z = 4.50$. *Astron. Astrophys.* **364**, L54–L61 (2000).
- Haislip, J. *et al.* GRB 050904: SOAR/PROMPT observations. *GCN Circ.* **3913** (2005).
- Haislip, J. *et al.* GRB 050904: possible high-redshift GRB. *GCN Circ.* **3914** (2005).
- Reichart, D. GRB 050904: environmental constraints. *GCN Circ.* **3915** (2005).
- Haislip, J. *et al.* GRB 050904: SOAR YJ and PROMPT Ic observations. *GCN Circ.* **3919** (2005).
- Kawai, N., Yamada, T., Kosugi, G., Hattori, T. & Aoki, K. GRB 050904: Subaru optical spectroscopy. *GCN Circ.* **3937** (2005).
- Cummings, J. *et al.* GRB050904: Swift-BAT detection of a probable burst. *GCN Circ.* **3910** (2005).
- Reichart, D. *et al.* PROMPT: panchromatic robotic optical monitoring and polarimetry telescopes. *Nuovo Cim.* **28**, 767–770 (2005).
- Fox, D. B. & Cenko, S. B. GRB050904: P60 observations. *GCN Circ.* **3912** (2005).
- Castro-Tirado, A. J. *et al.* Simultaneous and optical follow-up GRB observations by BOOTES. *Nuovo Cim.* **28**, 715–718 (2005).
- Jelinek, M. *et al.* GRB 050904: Bootes early R-band detection. *GCN Circ.* **3929** (2005).
- D'Avanzo, P. *et al.* GRB 050904: NIR object inside the XRT error box. *GCN Circ.* **3921** (2005).
- D'Avanzo, P. *et al.* GRB 050904: more VLT NIR observations. *GCN Circ.* **3930** (2005).
- Akerlof, C. *et al.* Observations of contemporaneous optical radiation from a gamma-ray burst. *Nature* **398**, 400–402 (1999).
- Li, W., Filippenko, A. V., Chornock, R. & Jha, S. The early light curve of the optical afterglow of GRB 021211. *Astrophys. J.* **586**, L9–L12 (2003).
- Fox, D. W. *et al.* Early optical emission from the γ -ray burst of 4 October 2002. *Nature* **422**, 284–286 (2003).
- Price, P. A. *et al.* The bright optical afterglow of the nearby γ -ray burst of 29 March 2003. *Nature* **423**, 844–847 (2003).
- Watson, D. *et al.* Outshining the Quasars at reionization: The X-ray spectrum and lightcurve of the redshift 6.29 γ -ray burst GRB 050904. *Astrophys. J.* (submitted).
- Reichart, D. E. Dust extinction curves and $\text{Ly}\alpha$ forest flux deficits for use in modeling gamma-ray burst afterglows and all other extragalactic point sources. *Astrophys. J.* **553**, 235–253 (2001).
- Spergel, D. N. *et al.* First-year Wilkinson microwave anisotropy probe (WMAP) observations: Determination of cosmological parameters. *Astrophys. J. Suppl. Ser.* **148**, 175–194 (2003).
- Klotz, A., Boer, M. & Atteia, J. L. GRB 050904: TAROT optical measurements. *GCN Circ.* **3917** (2005).
- Smith, J. A. *et al.* The $u'g'r'i'z'$ standard-star system. *Astron. J.* **123**, 2121–2144 (2002).
- Draine, B. T. Gamma-ray bursts in molecular clouds: H_2 absorption and fluorescence. *Astrophys. J.* **532**, 273–280 (2000).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Sakamoto, T. *et al.* GRB 050904 BAT refined analysis of complete data set. *GCN Circ.* **3938** (2005).
- Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* **497**, L17–L20 (1998).
- Chevalier, R. A. & Li, Z.-Y. Gamma-ray burst environments and progenitors. *Astrophys. J.* **520**, L29–L32 (1999).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).

Acknowledgements D.E.R. gratefully acknowledges support from NSF's MRI, CAREER, PREST and REU programmes, NASA's APRA, Swift GI and IDEAS programmes, and especially L. Goodman and H. Cox. D.E.R. also thanks W. Christiansen, B. Carney, and everyone who has worked to make SOAR a reality over the past 19 years. A.L. and N.T. thank B. Cavanagh and A. Adamson of the JAC for their speedy assistance in acquiring and reducing the UKIRT WFCAM data. A.J.C.-T. thanks INTA and CSIC for their support of BOOTES and AYA (including FEDER funds). R.H. and P.K. acknowledge support from GA AV CR and ESA PECS.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.E.R. (reichart@physics.unc.edu).

LETTERS

An optical spectrum of the afterglow of a γ -ray burst at a redshift of $z = 6.295$

N. Kawai¹, G. Kosugi², K. Aoki³, T. Yamada², T. Totani⁴, K. Ohta⁴, M. Iye², T. Hattori³, W. Aoki², H. Furusawa³, K. Hurley⁵, K. S. Kawabata⁶, N. Kobayashi⁷, Y. Komiyama³, Y. Mizumoto², K. Nomoto⁸, J. Noumaru³, R. Ogasawara², R. Sato¹, K. Sekiguchi³, Y. Shirasaki², M. Suzuki⁹, T. Takata², T. Tamagawa⁹, H. Terada³, J. Watanabe², Y. Yatsu¹ & A. Yoshida¹⁰

The prompt γ -ray emission from γ -ray bursts (GRBs) should be detectable out to distances of $z > 10$ (ref. 1), and should therefore provide an excellent probe of the evolution of cosmic star formation, reionization of the intergalactic medium, and the metal enrichment history of the Universe^{1–4}. Hitherto, the highest measured redshift for a GRB has been $z = 4.50$ (ref. 5). Here we report the optical spectrum of the afterglow of GRB 050904 obtained 3.4 days after the burst; the spectrum shows a clear continuum at the long-wavelength end of the spectrum with a sharp cut-off at around 9,000 Å due to Lyman α absorption at $z \approx 6.3$ (with a damping wing). A system of absorption lines of heavy elements at $z = 6.295 \pm 0.002$ was also detected, yielding the precise measurement of the redshift. The Si II fine-structure lines suggest a dense, metal-enriched environment around the progenitor of the GRB.

GRB 050904 was a long burst (duration $T_{90} = 225$ s) detected by the Swift γ -ray burst satellite on 4 September 2005, 01:51:44 UT (Universal time; refs 6, 7). Its position was immediately disseminated via the GRB Coordinates Network. Although the optical observations at the Palomar 60" telescope carried out 3.5 hours after the trigger did not reveal a new source with upper limits of $R > 20.8$ mag and $I > 19.7$ mag (ref. 8), a relatively bright near infrared source with $J \approx 17.5$ mag was detected three hours after the burst in the Swift X-ray telescope error circle⁹, which showed a temporal decay with an index of -1.20 , fully consistent with being a typical GRB afterglow. Analysis of the near-infrared colours, combined with the non-detection in the optical bands, led to the suggestion that the burst originated at a high redshift¹⁰, $5.3 < z < 9.0$. A refined photometric redshift of $z = 6.10^{+0.37}_{-0.12}$ was reported based on European Southern Observatory (ESO) Very Large Telescope (VLT) observations in the J , H , K and I bands¹¹.

We observed the field of GRB 050904 with the Faint Object Camera And Spectrograph (FOCAS)¹² on the 8.2-m Subaru Telescope on top of Mauna Kea, Hawaii, starting on the night of 6 September 2005. In the z' band image (600-s exposure, mid-epoch on 7 September, 8:04 UT), we detected the afterglow at $z'(\text{AB}) = 23.71 \pm 0.14$ mag, but we failed to detect it in the I_C band even with a longer exposure (900 s), which implied that the Lyman break should be present around $\lambda \approx 8,500$ – $9,000$ Å.

We then obtained a grism spectrum of the afterglow candidate, which exhibited a sharp cut-off at $\lambda \approx 9,000$ Å with strong depletion

of the continuum at shorter wavelengths, strikingly similar to the spectra of quasars¹³ at $z > 6$ except for the absence of a broad Ly α emission line. The emission is very weak in the wavelength range shorter than 8,900 Å. In particular, the flux is consistent with zero in the ranges 8,500–8,900 Å and 7,000–7,500 Å that extend shortward of the Ly α and Ly β wavelengths for $z \approx 6.3$. This is a clear signature of absorption by neutral hydrogen in the intergalactic medium at $z > 6$, and marks the first detection of a Gunn–Peterson trough¹⁴ from an object other than high- z quasars¹⁵. We also find weak emission features at $\sim 7,500$ – $8,300$ Å which are presumably leakage flux from the continuum emission that is also found in quasar spectra at similar redshifts.

At the longer-wavelength end of the spectrum is a flat continuum with a series of absorption lines, which we identify as Si II, O I, and C II lines at a common redshift of $z = 6.295 \pm 0.002$. We believe that this is the redshift of the GRB host galaxy, since no other absorption line system was observed at a redshift consistent with that of the Ly α break. This firm spectroscopic identification of the redshift breaks the previous record of GRB 000131 at $z = 4.50$ (ref. 5).

From a closer examination of the absorption lines, we find that they are not saturated and can be used to estimate the column densities of the heavy elements as shown in Table 1. Using the standard photospheric solar abundances¹⁶ we obtain the metallicity of these elements as $[C/H] = -2.4$, $[O/H] = -2.3$, $[Si/H] = -2.6$, and $[S/H] = -1.0$, where $\log[N_{\text{H I}} (\text{cm}^{-2})] \approx 21.3$ is assumed from a damped Ly α system model for the Ly α damping wing presented below. These values may not represent the typical abundances in the GRB host galaxy for several reasons. First, they are derived using only a single ionization state for each element. Depletion due to dust condensation may modify the Si abundance in particular. And second, the spatial distribution of the heavy elements may be significantly different from that of hydrogen. It is possible that the heavy elements are distributed only locally around the GRB source in a metal-enriched circumstellar shell, while the neutral hydrogen is distributed on a larger scale in or outside the host galaxy.

Further analysis of the Si lines allows us to constrain the scale of the absorbing metals. Using the equivalent width ratio of the fine-structure transition lines Si II* $\lambda = 1,264.7$ Å and Si II $\lambda = 1,260.4$ Å, the electron density n_e can be constrained¹⁷ as $\log[n_e (\text{cm}^{-3})] = 2.3 \pm 0.7$ for a reasonable temperature range of $10^3 \text{ K} < T < 10^5 \text{ K}$. Combined with the column density and the abundance of Si derived

¹Department of Physics, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan. ²National Astronomical Observatory of Japan, Osawa 2-21-1,

Mitaka, Tokyo 181-8588, Japan. ³Subaru Telescope, National Astronomical Observatory of Japan, 650 North A'ohoku Place, Hilo, Hawaii 96720, USA. ⁴Department of Astronomy, Kyoto University, Sakyo-ku, Kyoto, 606-8502, Japan. ⁵University of California at Berkeley, Space Sciences Laboratory, 7 Gauss Way, Berkeley, California 94720-7450, USA. ⁶Department of Physical Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. ⁷Institute of Astronomy, University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181-0015, Japan. ⁸Department of Astronomy, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. ⁹RIKEN (Institute of Physical and Chemical Research), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. ¹⁰Department of Physics, Aoyama Gakuin University, Sagami-hara, Kanagawa 229-8558, Japan.

Table 1 | Absorption lines detected in the spectrum of the optical afterglow of GRB 050904

Observed wavelength (Å)	Equivalent width (Å)	Column density log (cm ⁻²)	Line identification (element, Å)	Redshift, z
9,041.0 ± 0.8	4.5 ± 1.0	14.44 ^{+0.14} _{-0.16}	C IV, λ = 1,548.2 (N V, λ = 1,238.8)	4.840 ± 0.001 (6.298 ± 0.001)
9,055.9 ± 1.7	1.7 ± 1.0	14.21 ^{+0.25} _{-0.46}	C IV, λ = 1,550.8 (N V, λ = 1,242.8)	4.840 ± 0.001 (6.287 ± 0.001)
9,146.4 ± 1.8	3.8 ± 1.1	15.60 ^{+0.14} _{-0.17}	S II, λ = 1,253.8	6.295 ± 0.001
9,188.7 ± 2.6	6.1 ± 3.7	16.20 ^{+1.87} _{-0.92}	S II, λ = 1,259.5	6.295 ± 0.002
9,195.9 ± 1.2	8.3 ± 2.7	14.29 ^{+0.57} _{-0.39}	Si II, λ = 1,260.4	6.296 ± 0.001
9,225.8 ± 1.8	3.9 ± 1.1	13.63 ^{+0.13} _{-0.16}	Si II*, λ = 1,264.7	6.295 ± 0.001
9,499.1 ± 0.9	10.3 ± 1.9	15.85 ^{+0.39} _{-0.28}	O I, λ = 1,302.2	6.295 ± 0.001
9,737.2 ± 1.1	12.3 ± 2.4	15.41 ^{+0.30} _{-0.26}	C II, λ = 1,334.5	6.296 ± 0.001

The wavelengths and equivalent widths were derived by fitting a single gaussian. The column densities of lines were estimated by the standard curve of growth analysis²⁵. The equivalent widths in the table are observed ones, that is, not converted to the rest-frame. The quoted uncertainties are 1σ statistical errors. Most of the absorption lines are consistent with being at a single redshift of $z = 6.295 \pm 0.002$ within the statistical uncertainties. The absorption lines at $\lambda = 9,041.0$ Å and $9,055.9$ Å could be identified as N V $\lambda = 1,238.8$ Å, $\lambda = 1,242.8$ Å, if they are at a redshift similar to that of the other absorption lines. However, the derived redshifts of these two lines are significantly inconsistent with each other. Another possible identification is C IV $\lambda = 1,548.2$ Å, $\lambda = 1,550.8$ Å in an intervening system at $z = 4.840$, which we think is more likely.

above and assuming a hydrogen ionization fraction of 0.1, we obtain the physical depth of the absorbing system to be 0.4 pc with an uncertainty of a factor of ~ 10 , reflecting the statistical errors and the possible temperature range. These fine-structure lines have been found in GRB afterglow spectra^{18–20}, whereas they have never been clearly detected in quasar damped Ly α systems¹⁸. This is consistent with a local origin for the absorption such as a metal-enriched molecular cloud in the star-forming region or a dense metal-enriched shell nebula swept-up by a progenitor wind prior to the GRB onset suggested for GRB 021004 (refs 21, 22) and GRB 030226 (ref. 23). The column density of C II is also consistent with the calculation for a carbon-rich Wolf–Rayet wind²⁴.

As shown in Fig. 1, the Ly α cut-off exhibits the signature of a damping wing redward of the Ly α wavelength. To our knowledge, this is the first detection of significant neutral hydrogen absorption at

$z \gtrsim 6$, allowing us to explore the distribution of neutral hydrogen in the vicinity of a GRB, in the host galaxy, and/or in intergalactic space at very high redshifts. Such a study is difficult with high- z quasars owing to their enormous ultraviolet flux, which ionizes the surrounding environment, and owing to the presence of a strong Ly α emission line.

There are two possibilities for the nature of the absorber. It may be a damped Ly α system associated with the host galaxy, which has been observed in the afterglows of several GRBs at lower redshifts^{18–20}. The other possibility is the neutral hydrogen in the intergalactic medium (IGM) left over from the pre-reionization era². If the latter is the case, we can now measure the neutral fraction of the IGM at $z \gtrsim 6$, giving important information on the reionization history of the Universe. We find that the wing shape can be reproduced either by a damped Ly α system (see inset of Fig. 1) or by the IGM. A comprehensive

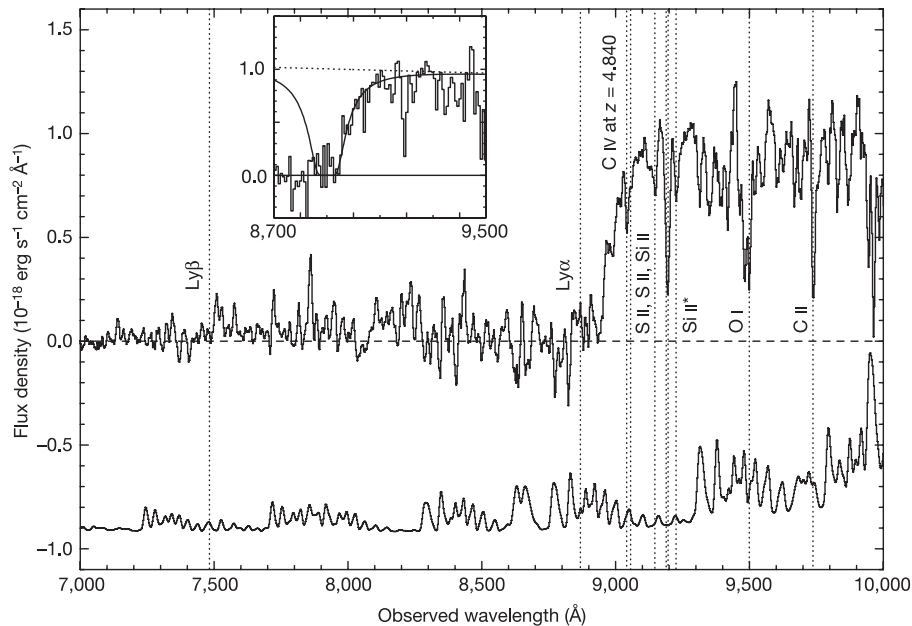


Figure 1 | The spectrum of the afterglow of GRB 050904. It covers the wavelengths 7,000–10,000 Å with a resolution $R = \lambda/\Delta\lambda \approx 1,000$ at 9,000 Å. It was taken with Subaru/FOCAS at mid-epoch on 7 September, 12:05 UT, 3.4 days after the burst, for a total exposure of 4.0 hours. The abscissa is the observed wavelength converted to that in vacuum. The spectrum is smoothed to a resolution $R = \lambda/\Delta\lambda \approx 600$ at 9,000 Å. The locations of the identified absorption lines (see Table 1) at $z = 4.840$ and $z = 6.295$, as well as the wavelengths of Ly α and Ly β at $z = 6.295$, are shown with vertical dotted lines. The one-sigma errors are shown with an offset of -1.0 at the bottom of

the panel. In the inset, the solid line shows a model for the damping wing of Ly α absorption with a neutral hydrogen column density $\log[N_{\text{H I}} (\text{cm}^{-2})] = 21.3$ at a redshift of $z = 6.3$, overlaid on the observed spectrum in the wavelength range of 8,700–9,500 Å. The dotted line indicates the unabsorbed continuum model following a power-law ($f_\nu \propto \nu^{-1}$) function as typically observed for GRB afterglows. We note that only the red wing is relevant to the fit, because the emission blueward of Ly α is absorbed by the IGM.

spectral fitting analysis is necessary to examine these possibilities, which is beyond the scope of this Letter.

With the detection of metal absorption lines, we have shown that GRBs are found in metal-enriched regions even at such an early phase of the Universe as $z > 6$. It is, therefore, possible that we would detect the metal absorption lines even from GRBs originating from the metal-free first-generation stars, because their environment may be self-polluted by pre-burst winds, as suggested by the present observation. We can expect to obtain afterglow spectra with much higher quality for GRBs at even higher redshifts in the immediate future, considering that Swift is constantly localizing faint GRBs, and that our spectrum was taken when the afterglow had faded by more than an order of magnitude since its first detection in the *J* band. Such future data will give us even better opportunities to probe the formation of stars and galaxies in the early Universe.

Received 20 September; accepted 2 December 2005.

1. Lamb, D. Q. & Reichart, D. E. Gamma-ray bursts as a probe of the very high redshift Universe. *Astrophys. J.* **536**, 1–18 (2000).
2. Miralda-Escude, J. Reionization of the Intergalactic Medium and the damping wing of the Gunn-Peterson trough. *Astrophys. J.* **501**, 15–22 (1998).
3. Totani, T. Cosmological gamma-ray bursts and evolution of galaxies. *Astrophys. J.* **486**, L71–L74 (1997).
4. Ciardi, B. & Loeb, A. Expected number and flux distribution of gamma-ray burst afterglows with high redshifts. *Astrophys. J.* **540**, 687–696 (2000).
5. Andersen, M. I. et al. VLT identification of the optical afterglow of the gamma-ray burst GRB 000131 at $z = 4.50$. *Astron. Astrophys.* **364**, L54–L61 (2000).
6. Cummings, J. et al. GRB050904: Swift-BAT detection of a probable burst. *GCN Circ.* **3910** (2005).
7. Sakamoto, T. et al. GRB 050904 BAT refined analysis of complete data set. *GCN Circ.* **3938** (2005).
8. Fox, D. B. & Cenko, S. B. GRB050904: P60 observations. *GCN Circ.* **3912** (2005).
9. Haislip, J. et al. GRB 050904: SOAR/PROMPT observations. *GCN Circ.* **3913** (2005).
10. Haislip, J. et al. GRB 050904: possible high-redshift GRB. *GCN Circ.* **3914** (2005).
11. Antonelli, L. A. et al. GRB 050904: photometric redshift. *GCN Circ.* **3924** (2005).
12. Kashikawa, N. et al. FOCAS: The Faint Object Camera and Spectrograph for the Subaru Telescope. *Publ. Astron. Soc. Jpn.* **54**, 819–832 (2002).
13. White, R. L., Becker, R. H., Fan, X. & Strauss, M. A. Probing the ionization state of the Universe at $z > 6$. *Astron. J.* **126**, 1–14 (2003).
14. Gunn, J. E. & Peterson, B. A. On the density of neutral hydrogen in intergalactic space. *Astrophys. J.* **142**, 1633–1641 (1965).
15. Becker, R. H. et al. Evidence for reionization at $z \sim 6$: Detection of a Gunn-Peterson trough in a $z = 6.28$ quasar. *Astron. J.* **122**, 2850–2857 (2001).
16. Grevesse, N. & Sauval, A. J. Standard solar composition. *Space Sci. Rev.* **85**, 161–174 (1998).
17. Silva, A. I. & Viegas, S. M. Physical conditions in quasi-stellar object absorbers from fine-structure absorption lines. *Mon. Not. R. Astron. Soc.* **329**, 135–148 (2002).
18. Vreeswijk, P. M. et al. The host of GRB 030323 at $z = 3.372$: A very high column density DLA system with a low metallicity. *Astron. Astrophys.* **419**, 927–940 (2004).
19. Chen, H.-W., Prochaska, J. X., Bloom, J. S. & Thompson, I. B. Echelle spectroscopy of a GRB afterglow at $z = 3.969$: A new probe of the interstellar and intergalactic media in the young Universe. *Astrophys. J. Lett.* (in the press); preprint at (<http://arXiv/astro-ph/0508270>) (2005).
20. Berger, E. et al. Spectroscopy of GRB 050505 at $z = 4.275$: A $\log N(\text{HI}) = 22.1$ DLA host galaxy and the nature of the progenitor. *Astrophys. J.* (submitted); preprint at (<http://arXiv/astro-ph/0511498>) (2005).
21. Mirabal, N. et al. GRB 021004: A possible shell nebula around a Wolf-Rayet star gamma-ray burst progenitor. *Astrophys. J.* **595**, 935–949 (2003).
22. Schaefer, B. E. et al. GRB 021004: A massive progenitor star surrounded by shells. *Astrophys. J.* **588**, 387–399 (2003).
23. Klose, S. et al. Probing a gamma-ray burst progenitor at a redshift of $z = 2$: A comprehensive observing campaign of the afterglow of GRB 030226. *Astron. J.* **128**, 1942–1954 (2004).
24. van Marle, A.-J., Langer, N. & Garcia-Segura, G. Constraints on gamma-ray burst and supernova progenitors through circumstellar absorption lines. *Astron. Astrophys.* (in the press); preprint at (<http://arXiv/astro-ph/0507659>) (2005).
25. Spitzer, L. *Physical Processes in the Interstellar Medium* 51 (Wiley-Interscience, New York, 1978).

Acknowledgements This work is based on data collected at the Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. We are grateful for support by the observatory. N.K. acknowledges support by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Tokyo Tech COE-21 programme 'Nanometer-scale Quantum Physics'. We thank S. Barthelmy for maintaining the GRB Coordinates Network, and the Swift team for providing rapid GRB localizations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.K. (nkawai@phys.titech.ac.jp).

Quasiparticle breakdown in a quantum spin liquid

Matthew B. Stone¹, Igor A. Zaliznyak², Tao Hong³, Collin L. Broholm^{3,4} & Daniel H. Reich³

Much of modern condensed matter physics is understood in terms of elementary excitations, or quasiparticles—fundamental quanta of energy and momentum^{1,2}. Various strongly interacting atomic systems are successfully treated as a collection of quasiparticles with weak or no interactions. However, there are interesting limitations to this description: in some systems the very existence of quasiparticles cannot be taken for granted. Like unstable elementary particles, quasiparticles cannot survive beyond a threshold where certain decay channels become allowed by conservation laws; their spectrum terminates at this threshold. Such quasiparticle breakdown was first predicted for an exotic state of matter—superfluid ⁴He at temperatures close to absolute zero, a quantum Bose liquid where zero-point atomic motion precludes crystallization^{1–4}. Here we show, using neutron scattering, that quasiparticle breakdown can also occur in a quantum magnet and, by implication, in other systems with Bose quasiparticles. We have measured spin excitations in a two-dimensional quantum magnet, piperazinium hexachlorodocuprate (PHCC)⁵, in which spin-1/2 copper ions form a non-magnetic quantum spin liquid, and find remarkable similarities with excitations in superfluid ⁴He. We observe a threshold momentum beyond which the quasiparticle peak merges with the two-quasiparticle continuum. It then acquires a finite energy width and becomes indistinguishable from a leading-edge singularity, so that excited states are no longer quasiparticles but occupy a wide band of energy. Our findings have important ramifications for understanding excitations with gapped spectra in many condensed matter systems, ranging from band insulators to high-transition-temperature superconductors⁶.

Although of all the elements only liquid helium fails to crystallize at temperature $T = 0$, quantum liquids are quite common in condensed matter. Metals host electron Fermi liquids, and superconductors contain Bose liquids of Cooper pairs. Trapped ultracold atoms can also form quantum liquids, and some remarkable new examples were recently identified in magnetic crystals^{5,7–10}. The organometallic material PHCC is an excellent physical realization of a quantum spin liquid (QSL) in a two-dimensional (2D) Heisenberg antiferromagnet (HAFM). Its Cu^{2+} spins are coupled through a complex network of orbital overlaps, and form an array of slightly skewed anisotropic spin-1/2 ladders¹⁰ in the crystalline a - c plane with highly frustrated super-exchange interactions⁵. The spin excitations in PHCC have a spectral gap $\Delta_s \approx 1$ meV and nearly isotropic 2D dispersion in the $(h0l)$ plane with a bandwidth slightly larger than Δ_s . In the absence of a magnetic field, only the short-range dynamic spin correlations typical of a liquid exist: the spin gap precludes long-range magnetic order down to $T = 0$. Here we explore magnetic excitations in PHCC via inelastic neutron scattering and compare the results with similar measurements in the quantum fluid ⁴He, emphasizing the effects where quasiparticle dispersion reaches the threshold for two-particle decay and interferes destructively with the continuum.

The properties of superfluid ⁴He (ref. 4) can be explained by considering Bose quasiparticles with a finite-energy minimum (an energy gap) in their spectrum^{1,2}. However, in a Bose quantum liquid, a spectral gap can produce an energy-momentum threshold where the quasiparticle description breaks down^{1–3}. Beyond this threshold, single-particle states are no longer approximate eigenstates of the hamiltonian and the quasiparticle spectrum terminates. Neutron scattering experiments in ⁴He indicate that the spectrum of quasiparticles (quanta of longitudinal sound waves also called phonons) ends when the phonon is able to decay into two ‘rotons’^{11–15}. These rotons are phonons with roughly quadratic dispersion that occur near the dispersion minimum, which is at energy $\Delta \approx 0.74$ meV and wavevector $Q \approx 2 \text{ \AA}^{-1}$ (compare Fig. 1, main panel). Spontaneous decays provide the only mechanism that destroys quasiparticles in ⁴He at $T = 0$. However, owing to the high density of two-roton states, this decay path is so effective that instead of acquiring a finite lifetime, the quasiparticles simply cease to exist. Specifically, the single-particle pole is absent in the Green’s function of ⁴He atoms for $Q > Q_c$ (where Q_c is a threshold wavevector), so that the quasiparticle spectrum does not continue beyond the threshold^{1–3}.

The excitation spectrum of superfluid ⁴He as probed by neutron scattering is shown in Fig. 1, main panel. One can see the roton minimum in the dispersion and the spectrum termination point at $Q_c \approx 2.6 \text{ \AA}^{-1}$. Near Q_c the phonon hybridizes with two-roton excitations, its dispersion flattens, and spectral weight is transferred

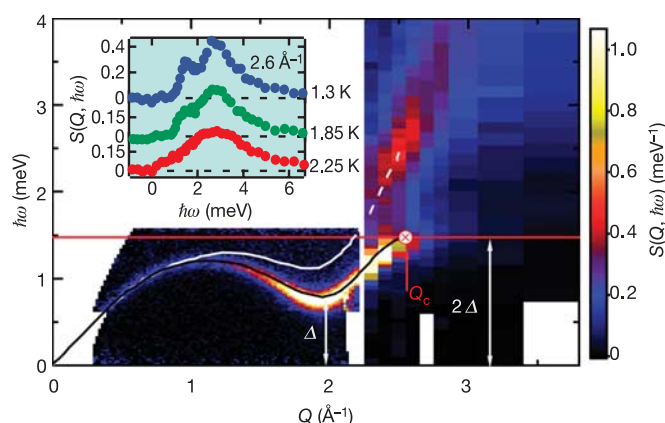


Figure 1 | Liquid helium excitation spectrum $S(Q, \hbar\omega)$ from inelastic neutron scattering measurements. Main panel, excitation spectrum in ⁴He for $1.5 \leq T \leq 1.8$ K. Data for wavevector $Q \geq 2.3 \text{ \AA}^{-1}$ are reproduced from ref. 13, data at smaller Q are from C.L.B. and S.-H. Lee, unpublished results. Solid black line, dispersion from ref. 13; red circle with cross, spectrum termination point at $Q = Q_c$ and $\hbar\omega = 2\Delta$. White line, Feynman–Cohen bare dispersion in absence of decays¹⁷; horizontal red line at $\hbar\omega = 2\Delta$, onset of two-roton states for $\hbar\omega \geq 2\Delta$. Inset, excitations near termination point, at $Q = 2.6 \text{ \AA}^{-1} \approx Q_c$, for several temperatures¹³.

¹Condensed Matter Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. ²Condensed Matter Physics and Material Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA. ³Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, Maryland 21218, USA.

⁴National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA.

to the multiparticle continuum^{13,15}. While a smeared maximum occurs at the leading edge of the continuum for $Q > Q_c$ and appears to continue the quasiparticle dispersion relation, it is instead ascribed to a two-roton bound state (resonance) resulting from roton–roton interactions^{15,16}. Decays modify the ‘bare’ Feynman–Cohen quasiparticle dispersion in ^4He (white line in Fig. 1, main panel)¹⁷. Instead of terminating where it reaches the energy 2Δ , the quasiparticle spectrum is suppressed to lower energies at $Q \leq Q_c$, approaching the threshold energy $\hbar\omega = 2\Delta$ horizontally³ (black line in Fig. 1, main panel).

The generality of the physics underlying quasiparticle breakdown in ^4He suggests that similar effects may occur in other quantum liquids. The quasiparticle instability in ^4He relies on the isotropic nature of the fluid: since the spectrum only depends on $|Q|$, the roton minimum produces a strong singularity in the density of states (DOS). For QSLs on a crystalline lattice, the DOS available for quasiparticle decays is enhanced by the absence of dispersion in certain directions that occurs in low-dimensional systems ($D < 3$) and in systems with competing interactions (frustration). Quasiparticle breakdown effects should thus be strongest in one-dimensional (1D) QSLs, such as spin-1 chains with a spectral gap¹⁸. Though the

term has not been used in this context, numerical work suggests that spectrum termination does occur in spin-1 HAFM spin chains^{19,20}. Its observation through neutron scattering, however, is hindered by small scattering cross-sections at the appropriate wavevectors. In the spin-1 chain system $\text{Ni}(\text{C}_2\text{H}_8\text{N}_2)_2\text{NO}_2\text{ClO}_4$ (NENP), scattering becomes undetectable when the single-particle excitation meets the non-interacting two-particle continuum²¹, owing to either decays or a vanishing structure factor. While transformation of magnetic excitations from well-defined quasiparticles to a continuum was observed in the quasi-1D spin-1 HAFM CsNiCl_3 , it is only seen as an onset of damping beyond a certain momentum threshold, well before the dispersion crosses the lower bound of the projected non-interacting two-particle continuum²², which may be a result of inter-chain interactions.

In contrast to the HAFM spin-1 chain, the structure factor of PHCC is favourable for probing the interaction of magnon quasiparticles with their two-particle continuum. Its effects, however, could be less pronounced because the 2D DOS singularities are weaker. Prior measurements examined magnetic excitations in PHCC below ~ 3 meV (ref. 5). Here we present data for energies $\hbar\omega \leq 7$ meV and for wavevectors along the $(1/2, 0, l)$ and $(h, 0, -1-h)$ directions, elucidating both single- and multiparticle excitations in this 2D QSL. Data shown in Fig. 2a and selected scans shown in Fig. 3 demonstrate clear similarities to the spectrum of superfluid ^4He . The one-magnon dispersion reaches the lower boundary of the two-magnon continuum, $\hbar\omega_{2m}(Q) = \min_q \{\hbar\omega(\mathbf{q}) + \hbar\omega(\mathbf{Q} - \mathbf{q})\}$, for $Q_c = (h_c, 0, -1-h_c)$ with

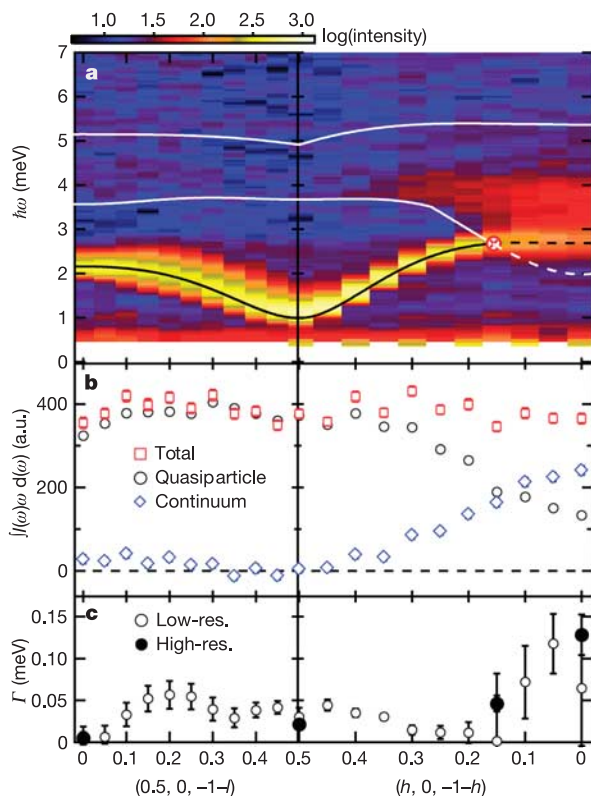


Figure 2 | Magnetic excitation spectrum at $T = 1.4$ K in PHCC.

a, Background-corrected intensity along the $(1/2, 0, -1-l)$ and $(h, 0, -1-h)$ directions. A $\delta\hbar\omega = 0.25$ meV running average was applied to each constant wavevector scan, retaining the actual point density of the acquired data. Black line, previously determined single-magnon dispersion⁵. White lines, bounds of two-magnon continuum calculated from this dispersion. Red circle with cross, the point where the single-particle dispersion relation intersects the lower bound of the two-particle continuum. **b**, First frequency moment of measured scattering intensity integrated over different energy ranges. Red squares (total), $0.8 \leq \hbar\omega \leq 5.5$ meV; black circles (quasiparticle), $0.8 \leq \hbar\omega \leq 3$ meV; blue diamonds (continuum), $3 \leq \hbar\omega \leq 5.5$ meV. **c**, Resolution-corrected half-width at half-maximum (HWHM) of the lower energy peak throughout the range of wavevector transfer for high resolution (solid points) and low resolution (open points) data. Error bars illustrate systematic error corresponding to 10% uncertainty in the neutron beam collimation used for resolution correction.

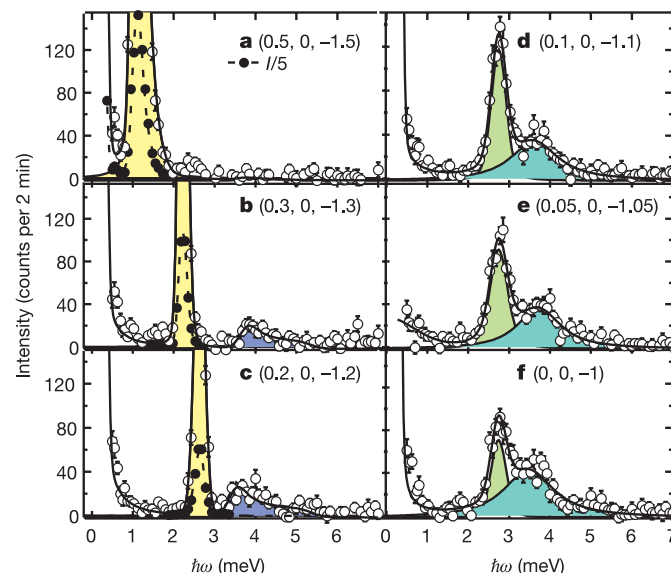


Figure 3 | Individual constant-wavevector scans of PHCC along the $(h, 0, -1-h)$ direction at $T = 1.4$ K.

Identical vertical scales emphasize variation in lineshape in the vicinity of h_c . **a–c**, Solid lines, fits to single resonant mode (yellow shaded region) plus a higher energy continuum excitation (blue shaded region) convolved with the instrumental resolution function. For wavevectors $h \geq 0.2$, higher energy excitations are well represented by a two-particle continuum of the form $I = A \frac{\Theta(\hbar\omega - \varepsilon_1(Q))\Theta(\varepsilon_2(Q) - \hbar\omega)}{\sqrt{(\hbar\omega)^2 - \varepsilon_1^2(Q)}}$ where Θ is a Heaviside step function and $\varepsilon_2(Q)$ is defined by the calculated upper boundary of the two-particle continuum (white line in Fig. 2a); $\varepsilon_1(Q)$ and A were refined by the least squares fitting. **d–f**, For $h \leq 0.15$ this description fails and the spectrum is fitted by two superimposed damped harmonic oscillator spectra, each in the form of a difference of two lorentzians whose HWHM Γ parameterizes damping, $I = \frac{\Gamma}{\pi} \left(\frac{1}{\Gamma^2 + (\hbar\omega - \hbar\omega_0)^2} - \frac{1}{\Gamma^2 + (\hbar\omega + \hbar\omega_0)^2} \right)$ (green shaded regions). The gaussian representing elastic incoherent nuclear scattering is also included at all wavevectors. Dashed lines and solid symbols in **a–c** show data on a one-fifth intensity (I) scale. Error bars show statistical uncertainty estimated as the square root of total neutron count measured at each point.

$h_c \approx 0.15$ near the magnetic Brillouin zone boundary. The first frequency moment²³ integrated over different ranges of energy transfer shown in Fig. 2b reveals how oscillator strength is transferred from the quasiparticle excitation to the multiparticle continuum, in analogy to what is observed in ^4He (ref. 13).

A change in the character of the excitation spectrum near h_c is also apparent in Fig. 3, which shows the energy-dependent magnetic scattering for wavevectors along the $(h, 0, -1-h)$ direction at $T \approx 1.4\text{ K} \ll \Delta_s$. For $h \geq 0.2$, Fig. 3a–c, there are two distinct contributions, a resolution-limited quasiparticle peak at lower energy and a broad feature with a sharp onset at higher energy, which we associate with the two-particle continuum. This continuum is well described by a square-root singularity above an energy threshold, typical for two-particle scattering governed by a diverging spectral density¹⁹. The threshold obtained from such data analysis is slightly higher than the calculated lower boundary of the two-magnon continuum (white line in Fig. 2a), and is close to the lowest energy of two-particle states involving gap mode magnons with a diverging DOS. Alternatively, the shift could indicate magnon repulsion.

For $h \leq 0.15$, the quasiparticle peak joins the continuum to form a complex spectral feature that extends from 2.5 to 4.5 meV (Fig. 3d–f). We parameterize this spectrum by the overlapping response of two damped harmonic oscillators. The onset of scattering occurs well above the lowest energy for two non-interacting magnons (dashed white line in Fig. 2a), which indicates significant interactions. While the lower energy peak that appears to continue the quasiparticle dispersion in PHCC carries more spectral weight than the corresponding resonance at $Q > Q_c$ in superfluid ^4He , it also has a measurable energy width as quantified in Fig. 2c. This demonstrates that a decay mechanism abruptly becomes accessible to the low energy excitations for $h \leq 0.15$. The width increases towards the Brillouin zone boundary, $h = 0$, where the peak at the leading edge can be described by a non-quasiparticle square root singularity as used for the continuum at $h \geq 0.2$, or as an unstable non-dispersive resonance below the continuum.

The temperature dependence of scattering in ^4He for Q between 2.4 and $2.6\text{ \AA}^{-1}$, Fig. 1 inset^{13,14}, provides additional evidence of quasiparticle spectrum breakdown. Data in PHCC for $Q = (0.15, 0, -1.15)$ where the one and two magnon states converge (Fig. 4e–h) similarly indicate that proximity to the two-particle continuum enhances thermal damping: the peak whose energy is approximately 20 K is already severely broadened at $T = 10\text{ K}$ (Fig. 4f). Its thermal broadening resembles that of the $Q = (0.5, 0, -1.5)$ gap mode, which is shown in Fig. 4a–d. This differs from observations in copper nitrate, a 1D QSL with weak dispersion where the one-magnon band lies well below the two-magnon continuum and decays cannot occur²⁴. Temperature-induced damping in that case is stronger for the lower-energy gap mode than for quasiparticles at the top of the dispersion curve; that is, heating mainly affects energy levels that become thermally populated. For PHCC, damping near the top and bottom of the band is governed by the same thermal population (Fig. 4a inset), consistent with the idea that high-energy excitations decay into gap-mode quasiparticles. As their thermal population increases, the probability of stimulated emission by the high-energy excitations also grows.

In summary, quasiparticle spectrum termination as seen in superfluid ^4He can also occur in other condensed matter systems, and in quantum magnets in particular. The dramatic changes observed in the spectrum of magnetic excitations in PHCC provide compelling evidence for such a phenomenon in the 2D QSL. The termination point is marked by rapid transfer of intensity from the magnon peak to the continuum at higher energies and by an abrupt appearance of damping. Although in PHCC the damped peak at the leading edge of magnetic scattering carries more intensity than the analogous peak in superfluid ^4He , the lineshape and temperature dependence of post-threshold excitations in these two very different quantum liquids are remarkably similar.

Quasiparticles are ubiquitous in nature, ranging from phonons, magnons, rotons^{1–3}, magnetorotons²⁵ and heavy electrons and holes in condensed matter physics to quasiparticles of the quark gluon plasma and various unstable particles and resonances in the standard model of particle physics²⁶. Rarely, however, do experiments offer as detailed a view of quasiparticle decay as the present results in a 2D organometallic spin liquid. Our findings show that an analysis of excitations in terms of quasiparticles with a well-defined dispersion relation can fail beyond a certain energy-momentum threshold where the quasiparticles break down. This has important implications for a variety of condensed matter systems, in particular for other QSLs such as lamellar copper oxide superconductors, where spin excitations above a gap are considered as possible mediators of electron pairing and high-temperature superconductivity²⁷.

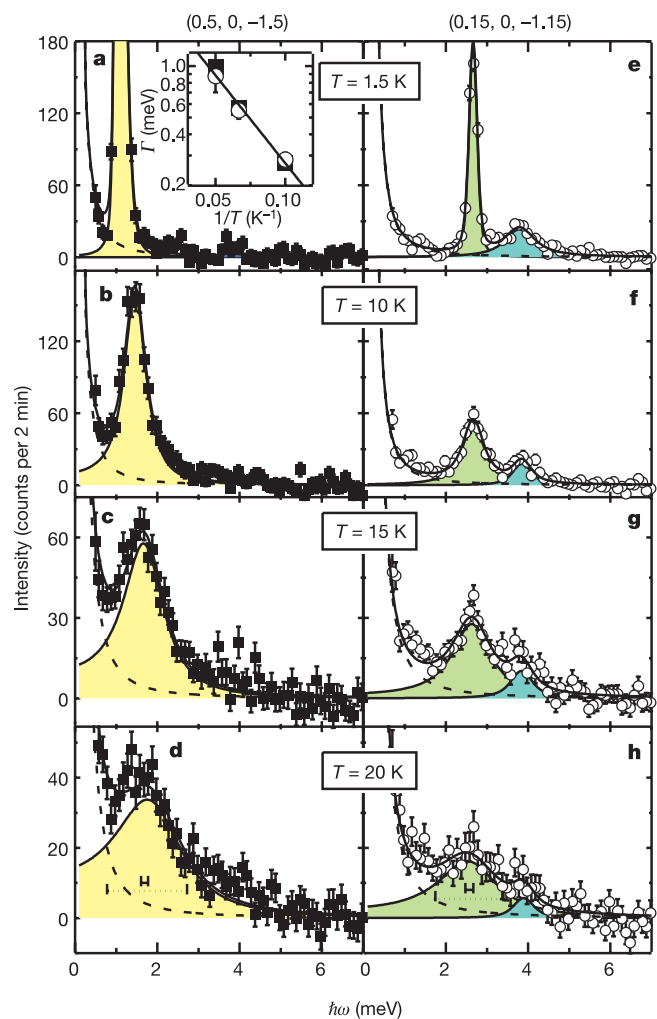


Figure 4 | Temperature dependent energy spectra for PHCC. Data are shown at $Q = (0.5, 0, -1.5)$ (a–d) and $Q = (0.15, 0, -1.15)$ (e–h). Solid lines for $T = 1.5\text{ K}$ in a and b are fits as described in Fig. 3. Solid lines for $T \geq 10\text{ K}$ are fits to the following response function satisfying a detailed balance constraint: $S(Q, \omega) = \frac{\Gamma}{\pi} \left(\frac{1}{\Gamma^2 + (\hbar\omega - \hbar\omega_0)^2} - \frac{1}{\Gamma^2 + (\hbar\omega + \hbar\omega_0)^2} \right) \frac{1}{1 - \exp(-\beta\hbar\omega)}$. The temperature dependence of the relaxation rate, Γ , for the lower energy peak at both wavevectors is shown in the inset to a. The line corresponds to exponentially activated behaviour with $\Delta = 2.0\text{ meV}$. Coloured areas below peaks indicate the assignment of different contributions to the spectra. Dashed lines indicate incoherent elastic nuclear scattering. Solid (dashed) horizontal bars in d and h indicate resolution (width of the low energy peak). Error bars show statistical uncertainty estimated as the square root of total neutron count measured at each point.

METHODS

Neutron scattering measurements of $(\text{C}_4\text{H}_{12}\text{N}_2)\text{Cu}_2\text{Cl}_6$ (PHCC) were performed using the SPINS cold neutron triple axis spectrometer at the NIST Center for Neutron Research. Four deuterated PHCC crystals⁵ with a total mass of 7.5 g were co-aligned to within 1°. Energy scans were acquired by varying the incident beam energy for fixed monitor counts in a low-efficiency detector between the pyrolytic graphite (PG (002)) monochromator and the sample. A 138' radial collimator was used between the sample and a horizontally focusing PG (002) analyser with an angular acceptance of 5° horizontally and 6° vertically, which was followed by a matching single-channel, high-efficiency cylindrical detector. A cooled Be filter was placed after the sample. Measurements in Fig. 4 employed an additional PG filter before the sample. Data in Figs 2 and 3 were acquired with 5 meV fixed final energy, and data in Fig. 4 at 3.7 meV fixed final energy. Projected full-width at half-maximum energy resolution of these configurations at $\hbar\omega = 0$ is 0.18 meV and 0.11 meV, respectively. A wavevector independent fast-neutron background was measured by shielding the analyser entrance with cadmium. A wavevector-dependent thermal neutron background arising predominantly from incoherent phonon scattering was measured at $T = 100$ K and scaled using the thermal detailed balance factor for use as a low-temperature non-magnetic background. These backgrounds were subtracted from all data presented.

Received 12 August 2005; accepted 10 January 2006.

- Landau, L. D. & Lifshitz, E. M. (eds) *Course of Theoretical Physics* Vol. 9, *Statistical Physics* Part 2 (by Lifshitz, E. M. & Pitaevskii, L. P.) 125–139 (Pergamon, Oxford, 1981).
- Abrikosov, A. A., Gorkov, L. P. & Dzyaloshinski, I. E. *Methods of Quantum Field Theory in Statistical Physics* 235–247 (Dover Publications, New York, 1975).
- Pitaevskii, L. P. Properties of the spectrum of elementary excitations near the disintegration threshold of the excitations. *Sov. Phys. JETP* **9**, 830–837 (1959).
- Kapitza, P. Viscosity of liquid helium below the λ -point. *Nature* **141**, 74 (1938).
- Stone, M. B., Zaliznyak, I., Reich, D. H. & Broholm, C. Frustration-induced two dimensional quantum disordered phase in piperazinium hexachlorodiprurate. *Phys. Rev. B* **64**, 144405 (2001).
- Anderson, P. W. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science* **235**, 1196–1198 (1987).
- Xu, G. et al. Holes in a quantum spin liquid. *Science* **289**, 419–422 (2000).
- Rüegg, Ch. et al. Bose-Einstein condensation of the triplet states in the magnetic insulator TiCuCl_3 . *Nature* **423**, 62–65 (2003).
- Stone, M. B. et al. Frustrated three-dimensional quantum spin liquid in CuHpCl . *Phys. Rev. B* **65**, 064423 (2002).
- Gopalan, S., Rice, T. M. & Sigrist, M. Spin ladders with spin gaps: A description of a class of cuprates. *Phys. Rev. B* **49**, 8901–8910 (1994).
- Woods, A. D. B. & Cowley, R. A. Structure and excitations of liquid helium. *Rep. Prog. Phys.* **36**, 1135–1231 (1973).
- Smith, A. J., Cowley, R. A., Woods, A. D. B., Stirling, W. G. & Martel, P. Roton-roton interactions and excitations in superfluid helium at large wavevectors. *J. Phys. C* **10**, 543–553 (1977).
- Fåk, B. & Bossy, J. Temperature dependence of $S(Q, E)$ in liquid ^4He beyond the roton. *J. Low Temp. Phys.* **112**, 1–18 (1998).
- Montfrooij, W. & Svensson, E. C. Superfluid ^4He : just a very cold liquid? *J. Low Temp. Phys.* **121**, 293–302 (2000).
- Pistoiesi, F. Theory and data analysis for excitations in liquid ^4He beyond the roton minimum. *Phys. Rev. Lett.* **81**, 397–400 (1998).
- Ruvalds, J. & Zawadowski, A. Theory of structure in the superfluid helium spectrum considering roton-roton resonances. *Phys. Rev. Lett.* **25**, 333–337 (1970).
- Feynman, R. P. & Cohen, M. Energy spectrum of the excitations in liquid helium. *Phys. Rev.* **102**, 1189–1204 (1956).
- Giamarchi, T. *Quantum Physics in One Dimension* (International Series of Monographs on Physics, Oxford Univ. Press, Oxford, UK, 2005).
- Meshkov, S. V. Monte-Carlo study of quantum spin chains. *Phys. Rev. B* **48**, 6167–6178 (1993).
- Regnault, L.-P., Zaliznyak, I. A. & Meshkov, S. V. Thermodynamic properties of the Haldane spin chain: statistical model for the elementary excitations. *J. Phys. Condens. Matter* **5**, L677–L683 (1993).
- Ma, S., Broholm, C., Reich, D. H., Sternlieb, B. J. & Erwin, R. W. Dominance of long-lived excitations in the antiferromagnetic spin-1 chain NENP. *Phys. Rev. Lett.* **69**, 3571–3574 (1992).
- Zaliznyak, I. A., Lee, S.-H. & Petrov, S. V. Continuum in the spin-excitation spectrum of a Haldane chain observed by neutron scattering in CsNiCl_3 . *Phys. Rev. Lett.* **87**, 017202 (2001).
- Hohenberg, P. C. & Brinkmann, W. F. Sum rules for the frequency spectrum of linear magnetic chains. *Phys. Rev. B* **10**, 128–131 (1974).
- Xu, G., Broholm, C., Reich, D. H. & Adams, M. A. Triplet waves in a quantum spin liquid. *Phys. Rev. Lett.* **84**, 4465–4468 (2000).
- Girvin, S. M., MacDonald, A. H. & Platzman, P. M. Magneto-roton theory of collective excitations in the fractional quantum Hall-effect. *Phys. Rev. B* **33**, 2481–2494 (1986).
- Cottingham, W. N. & Greenwood, D. A. *An Introduction to the Standard Model of Particle Physics* (Cambridge Univ. Press, Cambridge, UK, 1998).
- Tranquada, J. M. et al. Quantum magnetic excitations from stripes in copper oxide superconductors. *Nature* **429**, 534–538 (2004).

Acknowledgements We acknowledge discussions with L. Passell, J. Tranquada, A. Abanov, M. Zhitomirsky, A. Tsvetlik, A. Chitov and M. Swartz. Work at BNL and ORNL was supported by the US Department of Energy. Work on SPINS and at JHU was supported by the US National Science Foundation. We are grateful to B. Fåk for permission to reproduce the ^4He results in Fig. 1 and to A. Zaliznyak for help with the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.A.Z. (zaliznyak@bnl.gov).

Adsorption-induced scission of carbon-carbon bonds

Sergei S. Sheiko¹, Frank C. Sun¹, Adrian Randall¹, David Shirvanyants¹, Michael Rubinstein¹, Hyung-il Lee² & Krzysztof Matyjaszewski²

Covalent carbon-carbon bonds are hard to break. Their strength is evident in the hardness of diamonds^{1,2} and tensile strength of polymeric fibres³⁻⁶; on the single-molecule level, it manifests itself in the need for forces of several nanonewtons to extend and mechanically rupture one bond. Such forces have been generated using extensional flow⁷⁻⁹, ultrasonic irradiation¹⁰, receding meniscus¹¹ and by directly stretching a single molecule with nano-probes¹²⁻¹⁶. Here we show that simple adsorption of brush-like macromolecules with long side chains on a substrate can induce not only conformational deformations¹⁷, but also spontaneous rupture of covalent bonds in the macromolecular backbone. We attribute this behaviour to the fact that the attractive interaction between the side chains and the substrate is maximized by the spreading of the side chains, which in turn induces tension along the polymer backbone. Provided the side-chain densities and substrate interaction are sufficiently high, the tension generated will be strong enough to rupture covalent carbon-carbon bonds. We expect similar adsorption-induced backbone scission to occur for all macromolecules with highly branched architectures, such as brushes and dendrimers. This behaviour needs to be considered when designing surface-targeted macromolecules of this type—either to avoid undesired degradation, or to ensure rupture at predetermined macromolecular sites.

A series of brush-like macromolecules with the same number average degree of polymerization of a poly(2-hydroxyethyl methacrylate) backbone, $N_n = 2,150 \pm 100$, and different degrees of polymerization of poly(*n*-butyl acrylate) (pBA) side chains ranging from $n = 12 \pm 1$ to $n = 140 \pm 12$ were synthesized by atom transfer radical polymerization (see 'Polymer Characterization' in the Methods)¹⁸. Owing to the high grafting density, the side chains

repel each other and thereby stretch the backbone into an extended conformation. Placing these macromolecules on a surface enhances the steric repulsion between the side chains, which results in both an extension of the polymer backbone and an increase of the persistence length.

The effect is illustrated in Fig. 1, which shows atomic force microscopy (AFM) micrographs of monolayers of pBA brushes with short (Fig. 1a) and long side chains (Fig. 1b). Measurements on both types of molecules yielded a number average contour length per monomeric unit of the backbone of $l = L_n/N_n = 0.23 \pm 0.02$ nm (see 'Atomic Force Microscopy' in Methods), which is close to $l_0 = 0.25$ nm, the length of the tetrahedral C-C-C section. This means that even for short side chains ($n = 12$), the backbone is already fully extended and adopts an all-*trans* conformation. As the side chains become longer, we observe global straightening of the backbone reflected in the increase of the persistence length (Fig. 1c).

Chain extension requires a substantial amount of force, which we estimate using simple spreading arguments (Fig. 2). Just as in normal liquids, the polymeric side chains spread to cover the higher-energy substrate and thus stretch the macromolecule in all directions. Unlike conventional liquids, however, the spreading of the side chains is constrained by their connection to the backbone. To maximize the number of side chains that adsorb to the substrate, the backbone needs to extend; but even when it is fully elongated, about 50% of the side chains are still not fully in contact with the substrate. In this situation, the attraction of the side chains for the surface causes the polymer backbone to extend beyond its physical limit. Here it is important that the tension imposed by the surface attraction is unevenly distributed over the covalent bonds of the molecular skeleton.

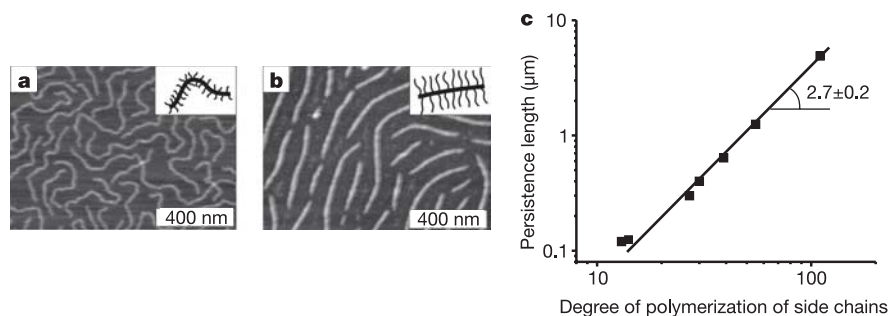


Figure 1 | Conformational response of pBA brush-like macromolecules to adsorption on mica. The conformation of the macromolecules is visualized by AFM, with the light threads in the height images shown in **a** and **b** corresponding to the backbones. The areas between threads are covered by side chains, which cannot be visualized at this scale. With increasing side-

chain length, molecules change from a fairly flexible conformation for $n = 12$ (shown in **a**) to a rod-like conformation for $n = 130$ (shown in **b**). **c**, The persistence length l_p of the adsorbed macromolecules was determined from the statistical analysis of the backbone curvature. It is found to increase with the side chain length as $l_p \propto n^{2.7}$.

¹Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290, USA. ²Center for Macromolecular Engineering, Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.

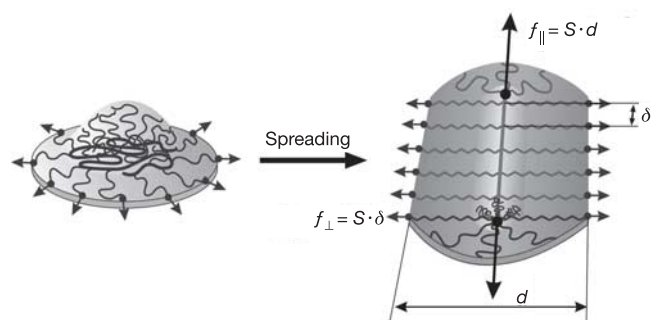


Figure 2 | Schematic of the spreading of a brush-like macromolecule on an attractive substrate. After adsorption, the macromolecule spreads to increase the number of monomeric contacts with the substrate. The brush-like architecture imposes constraints on the spreading process making it anisotropic and leading to extension of the backbone. Along the brush axis, the wetting-induced tensile force $f \approx S \cdot d$ is supported almost entirely by the covalently linked backbone, where S is the spreading coefficient and d is the brush width. In the direction perpendicular to the backbone, the force is evenly distributed over many side chains, each bearing a $f \approx S \cdot \delta$ tensile force, where δ is the distance between the neighbouring side chains.

As shown in Fig. 2, along the brush axis, a major fraction of the wetting-induced tensile force is carried by the backbone; while in the perpendicular direction, the tension is evenly distributed over many side chains. The force at the backbone is estimated as $f \approx S \cdot d$, where S is the spreading coefficient and d is the width of adsorbed brush macromolecules (Fig. 2). Here, we consider only the dominant term in S : that is, the difference between the surface free energies of substrate–gas, liquid–gas, and substrate–liquid–gas interfaces ($S = \gamma_s - \gamma_l - \gamma_{sl}$). Previous measurements for the substrates that were used in this study found $S \approx 20 \text{ mN m}^{-1}$ on graphite¹⁹ and water/alcohol mixtures¹⁷. Therefore, a brush macromolecule with short side chains ($n = 12$) and a width of $d = 11 \text{ nm}$ (ref. 20), is

capable of generating a force of approximately 220 pN on either of these two substrates. This exceeds the typical range of tensile forces of 10–100 pN reported for stretching of individual polymer chains²¹.

According to these arguments, the force value is proportional to the molecule's width and also depends on the surface energy of the substrate. We therefore synthesized pBA brushes with longer side chains ($n = 140$) that would lead to a width of $d = 130 \text{ nm}$ (refs 20, 22); this should result in a tensile force of about 2.6 nN and allow us to challenge the carbon–carbon bonds in the backbone⁹. The molecules were adsorbed on the surface of mica, graphite, silicon wafers and a range of water/propanol mixtures. Whereas molecules on solid substrates could be directly imaged by AFM, the liquid-supported films were first transferred onto a solid substrate using the Langmuir–Blodgett technique and then scanned by AFM (see Methods).

Figure 3a shows a series of AFM images obtained for different incubation times on the water/propanol (99.8/0.2 wt/wt%) substrate, which has a surface energy of $\gamma_s = 69 \pm 1 \text{ mN m}^{-1}$ and a spreading parameter of $S = 21 \pm 2 \text{ mN m}^{-1}$, where the experimental errors are determined by the precision of the Wilhelmy plate method (see 'Langmuir–Blodgett monolayers' in the Methods). As the time spent on the substrate increases, the molecules get progressively shorter while their number density (number of molecules per unit area) correspondingly increases; this suggests scission of the backbone (Fig. 3b). The cumulative length of molecules per unit mass of the material was measured as $\Lambda = \frac{\sum L_i \cdot n_i}{\sigma \times A}$, where n_i is the number of molecules of length L_i within a substrate area A and σ is the Langmuir–Blodgett-controlled mass per unit area of the monolayer. As shown in Fig. 3c, the cumulative length remains approximately constant for different exposure times, supporting the idea that chain scission occurs. Similar observations were made on other substrates (Supplementary Fig. 1). However, we focused on experiments using liquid substrates because they allow gradual variation of the surface energy simply by mixing two different liquids, and because they

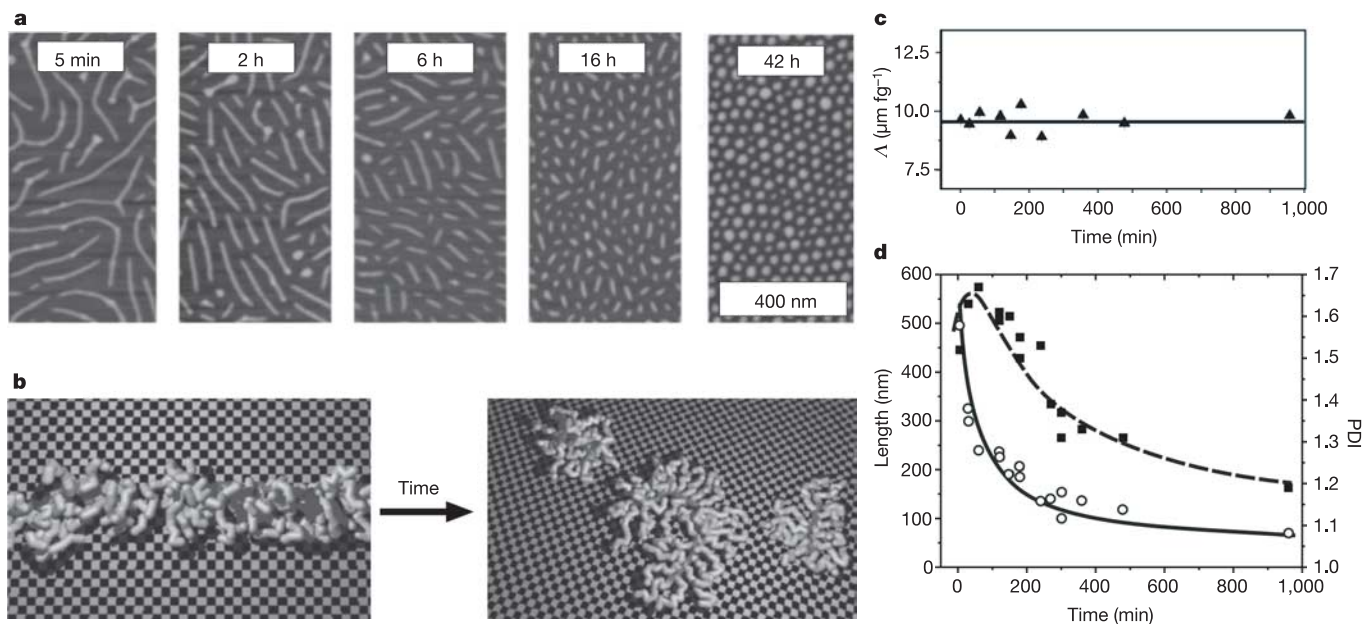


Figure 3 | Adsorption-induced degradation of macromolecules. **a**, The molecular degradation of brush-like macromolecules with long side chains ($n = 140$) on mica was monitored using AFM height imaging after each sample was exposed for different time periods (as indicated in the images) to a water/propanol (99.8/0.2 wt/wt%) substrate. **b**, Schematics of an adsorbed macromolecule (left) which undergoes spontaneous scission of the covalent backbone (right). Side chains are shown in light grey, the backbone in dark grey. **c**, The cumulative length per unit mass, measured within an area of

$A = 25 \mu\text{m}^2$ at a constant mass density of $\sigma = 0.08 \mu\text{g cm}^{-2}$, was found to stay at an approximately constant value of $\Lambda = 9.6 \pm 0.5 \mu\text{m fg}^{-1}$ throughout the scission process. **d**, The number average contour lengths measured after different exposure times t (white circles) are fitted according to $\frac{1}{L-L_\infty} = \frac{1}{L-L_0} + \frac{kt}{L_\infty}$, using experimental values for L_0 and L_∞ and a fitted value for k of $2.3 \times 10^{-5} \text{ s}^{-1}$ (solid line). The experimentally determined polydispersity index $\text{PDI} = L_w/L_n$ (black squares) shows good agreement with the computer simulation results of Fig. 4 (dashed line).

facilitate rapid equilibration of the monolayer structure. Both factors ensure reproducibility of the kinetics study discussed below.

Figure 3d shows the characteristic decay of the average molecular length with increasing exposure time of the macromolecules to the water/propanol substrate. To analyse the kinetics of the scission process we assume that the bond scission occurs as a first-order reaction: $B = B_0 e^{-kt}$, where B is the total number of covalent bonds in all backbones within a unit area of the substrate, B_0 is the initial number of bonds at $t = 0$, and k is the rate constant. Because the cumulative length is conserved, we can obtain the number average contour length L from $\frac{1}{L-L_\infty} = \frac{1}{L_0-L_\infty} + \frac{kt}{L_\infty}$, where $L_0 = 496 \pm 18$ nm is the initial contour length measured by AFM at $t = 0$ and $L_\infty = 40 \pm 3$ nm is the length of the shortest molecule observed during the scission process (see 'Atomic Force Microscopy' in the Methods). Fitting the experimental data to this equation using k as a fitting parameter yielded $k = 2.3 \times 10^{-5} \text{ s}^{-1}$. That we did not observe molecules shorter than 40 nm even at very long exposure times is because the brush molecules with short backbones adopted star-like morphologies. This ensures that the side chains have more space to spread out and eases tension at the backbone. The reduction of tension prevents further scission, so the above rate equations are applied only at $L \geq L_\infty$.

The scission process seems random, which suggests a uniform distribution of tension along the backbone. We probe this assumption by analysing the length distribution of the system throughout the scission process. As shown in Fig. 3d, the polydispersity index, $\text{PDI} = L_w/L_n$, initially increases and then decays, where L_w and L_n are the weight and number average lengths of adsorbed macromolecules, respectively. This is consistent with random cleavage of backbone C–C bonds, which initially increases the length polydispersity and then results in an almost monodisperse system as

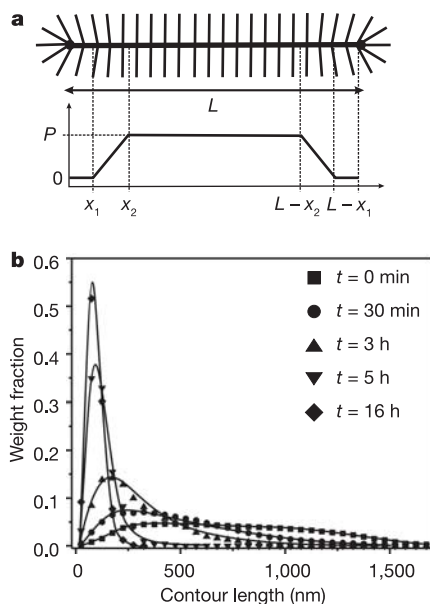


Figure 4 | Computer simulation of the scission process. **a**, The computer model assumes a constant scission probability P along most of the backbone; at the ends, P decays linearly to zero from $x_2 = 120$ nm to $x_1 = 40$ nm. This ensures the scission process stops at the experimentally observed $L_\infty = 40$ nm. **b**, Length distributions obtained by computer simulation for different time intervals t of the scission process (solid lines). The simulated distributions show good agreement with the distributions (data points) obtained by AFM on the same polymer/substrate system as used to obtain the images shown in Fig. 3a. The distributions are presented as the weight fraction of polymer chains of a certain number average contour length with a resolution (bin size) of 50 nm. The initial distribution function exactly corresponds to a realistic ensemble of 2,450 molecules acquired by AFM at $t = 0$ with $L_n = 496$ nm and $\text{PDI} = 1.52$.

the macromolecules gradually convert into short brush molecules that can no longer undergo scission. This behaviour can be simulated using a simple model wherein the probability P of bond scission occurring at any point along the backbone, except at the ends, is the same (Fig. 4a). Solid lines in Fig. 4b depict length distributions obtained at different durations of the computer-simulated scission process compare favourably with the corresponding length distributions measured by AFM (data points in Fig. 4b), giving good agreement between the modelled and experimentally measured polydispersity index. The simulated scission process eventually stops when all molecules become shorter than 80 nm, that is, in the range from $x_1 = 40$ nm and $2x_1 = 80$ nm.

Experiments are also being conducted to verify the effects of the substrate surface energy and the side-chain length on scission. As might be expected, preliminary findings show that backbone scission is very sensitive to small variations in both parameters. If surface energy is decreased to below 60 mN m^{-1} by adding more propanol to the water/propanol mixture used as substrate, molecules with long side chains ($n = 140$) that readily break on a 99.8/0.2 wt/wt% water/propanol surface ($\gamma = 69 \text{ mN m}^{-1}$) remain intact. Sharp retardation of the scission process was also observed upon shortening of the side chains: when using the same substrate (that is, a 99.8/0.2 wt/wt% water/propanol mixture) but pBA brush-like molecules with $n = 130$ ($d = 120$) instead of $n = 140$ does not lead to any noticeable shortening within reasonable experimental times (for example, days). However, we found that these molecules break on graphite, which has a slightly higher surface energy and a spreading parameter (Supplementary Fig. 1).

The essential feature of the bond scission observed here is that it occurs spontaneously upon adsorption onto a substrate. Linear and weakly branched polymer chains are obviously not at risk of chemical degradation upon surface adsorption; but all highly branched macromolecules that physically cannot allow all their monomeric units to interact with a substrate will be susceptible. In such cases, the load imposed by the adsorption forces is unevenly distributed over different structural elements of the molecular architecture according to the branching topology. In the system we studied, tension is concentrated along the backbone of molecular brushes and can be enough to break covalent carbon–carbon bonds. In the case of regular dendrimers, tension will focus at the covalent bonds near the principal branching centre of the dendrimers and, if the adsorption forces are strong enough, can cause the dendrimers to break. This essentially geometric effect is closely related to the observation that dendrimer polymerization stops above a certain generation, owing to the overcrowded molecular volume²³. These steric constraints can be eased by increasing the length of the spacer between branches in dendrimers, and between side chains in cylindrical brushes.

However, these structural modifications that make the branched structure looser also increase the 'footprint' of the adsorbed macromolecule, which in turn leads to a greater tensile force. Thus, with the current pursuit of new macro- and supramolecular materials that are specifically tailored for various surface applications, the surface-induced scission of covalent bonds will need to be considered carefully when designing complex molecular architectures. But in addition to emphasizing the need for designing 'stress-free' macromolecules for some applications, the phenomenon described here also opens up intriguing opportunities for deliberately designing architectures that break at pre-defined sites.

METHODS

Polymer characterization. Average molecular weights and molecular weight distribution of brush-like macromolecules were measured by gel permeation chromatography (GPC) equipped with Waters microstyragel columns (pore sizes 10^5 , 10^4 and 10^3 Å) and three detection systems: a differential refractometer (Waters Model 410), multi-angle laser light-scattering (MALLS) detector (Wyatt, DAWN EOS), and a differential viscometer (WGE Dr. Bures, η -1001).

In addition, we used a newly developed approach based on a combination of AFM and Langmuir–Blodgett techniques²⁴. This combination of methods ensured relative experimental errors in determining the polymerization degrees of the backbone and side chains below 5% and 10%, respectively.

Langmuir–Blodgett monolayers. To study the kinetics of the scission process, brush-like macromolecules with pBA side chains were adsorbed onto a surface of a water/propanol (99.8/0.2 wt/wt%) substrate. Propanol was chosen for its low surface energy and because its vapour pressure is nearly equivalent to that of water. This was necessary for long incubation times so that any subphase evaporation would lead to a minimal change in the surface energy. The evaporation of the subphase was closely monitored and controlled in an environmental chamber. For AFM analysis, the monolayer films were transferred onto a mica substrate at a controlled transfer ratio of 0.98, using the Langmuir–Blodgett technique. The surface tension of the substrate and the corresponding spreading parameter were measured by the Wilhelmy plate method.

Atomic force microscopy. Topographic images of individual molecules were collected using an atomic force microscope (Veeco Metrology Group) in tapping mode. We used silicon cantilevers (Mikromasch-USA) with a resonance frequency of about 140 kHz and a spring constant of about 5 N m. The radius of the probe was less than 10 nm. The analysis of digital images was performed using a custom software program (PEN) developed in-house and available from S.S.S. The program identifies the molecular contour and is capable of determining the contour length, the end-to-end distance, and the curvature distribution, all required for evaluation of the persistence length. For every sample, about ten images of about 300 molecules, that is, a total of 3,000 molecules were measured to ensure a relative standard error below 4% and an experimental error below 5% of the persistence length (Fig. 1), contour length, and polydispersity index (Fig. 3) measurements.

Received 17 August 2005; accepted 5 January 2006.

- Field, J. E. Mechanical and physical properties of diamond. *Inst. Phys. Conf. Ser.* **75**, 181–205 (1986).
- Derrington, A. Why diamonds top a hard league. *The Financial Times* n33034, section 3, pWFT2(1) (13/14 July 1996).
- Smith, P. & Lemstra, P. J. Tensile strength of highly oriented polyethylene. *J. Polym. Sci. Polym. Phys. Ed.* **19**, 1007–1009 (1981).
- Porter, R. S. & Kanamoto, T. A review on the tensile strength of polyethylene fibers. *Polym. Eng. Sci.* **34**, 266–268 (1994).
- Salvetat, J.-P. *et al.* Elastic and shear moduli of single-walled carbon nanotube ropes. *Phys. Rev. Lett.* **82**, 944–947 (1999).
- Yu, M.-F., Files, B. S., Arepalli, S. & Ruoff, R. S. Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties. *Phys. Rev. Lett.* **84**, 5552–5555 (2000).
- De Gennes, P. G. Coil stretch transition of dilute flexible polymers under ultrahigh velocity gradients. *J. Chem. Phys.* **60**, 5030–5042 (1974).
- Harrington, R. E. & Zimm, B. H. Degradation of polymers by controlled hydrodynamic shear. *J. Phys. Chem.* **69**, 161–176 (1965).
- Odell, J. A., Keller, A. & Rabin, Y. Flow induced scission of isolated macromolecules. *J. Chem. Phys.* **88**, 4022–4028 (1988).
- Basedow, A. M. & Ebert, K. H. Ultrasonic degradation of polymers in solution. *Adv. Polym. Sci.* **22**, 83–148 (1977).
- Bensimon, A. *et al.* Alignment and sensitive detection of DNA by a moving interface. *Science* **265**, 2096–2098 (1994).
- Kishino, A. & Yanagida, T. Force measurements by micromanipulation of single actin filament by glass needles. *Nature* **334**, 74–76 (1988).
- Evans, E. Probing the relation between force—lifetime—and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105–128 (2001).
- Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
- Cluzel, P. *et al.* DNA: An extensible molecule. *Science* **271**, 792–794 (1996).
- Mehta, A. D., Rief, M., Spudich, J. A., Smith, D. A. & Simmons, R. M. Single-molecule biomechanics with optical methods. *Science* **283**, 1689–1695 (1999).
- Sun, F., Sheiko, S. S., Moller, M., Beers, K. & Matyjaszewski, K. Conformational switching of molecular brushes in response to the energy of interaction with the substrate. *J. Phys. Chem. A* **108**, 9682–9686 (2004).
- Matyjaszewski, K. & Xia, J. Atom transfer radical polymerization. *Chem. Rev.* **101**, 2921–2990 (2001).
- Xu, H. *et al.* Molecular motion in a spreading precursor film. *Phys. Rev. Lett.* **93**, 206103 (2004).
- Neugebauer, D., Sumerlin, B., Matyjaszewski, K., Goodhart, B. & Sheiko, S. S. How dense are cylindrical brushes grafted from a multifunctional macroinitiator? *Polymer* **45**, 8173–8179 (2004).
- Austin, R. H., Brody, J. P., Cox, E. C., Duke, T. & Volkmuth, W. Stretch genes. *Phys. Today* 32–38 (February 1997).
- Sheiko, S. S. & Möller, M. Visualization of molecules—a first step to manipulation and controlled response. *Chem. Rev.* **101**, 4099–4123 (2001).
- Boris, D. & Rubinstein, M. A self-consistent mean field model of a starburst dendrimer: dense core vs. dense shell. *Macromolecules* **29**, 7251–7260 (1996).
- Sheiko, S. S. *et al.* Measuring molecular weight by atomic force microscopy. *J. Am. Chem. Soc.* **125**, 6725–6728 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. T. Samulski and I. LaRue for their interest and discussions. This work was supported by grants from the National Science Foundation.

Author Contributions All authors contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.S.S. (sergei@email.unc.edu).

Photosensitized reduction of nitrogen dioxide on humic acid as a source of nitrous acid

Konrad Stemmler¹, Markus Ammann¹, Chantal Donders^{1,3}, Jörg Kleffmann² & Christian George⁴

Nitrous acid is a significant photochemical precursor of the hydroxyl radical^{1–13}, the key oxidant in the degradation of most air pollutants in the troposphere. The sources of nitrous acid in the troposphere, however, are still poorly understood. Recent atmospheric measurements^{7,10–17} revealed a strongly enhanced formation of nitrous acid during daytime via unknown mechanisms. Here we expose humic acid films to nitrogen dioxide in an irradiated tubular gas flow reactor and find that reduction of nitrogen dioxide on light-activated humic acids is an important source of gaseous nitrous acid. Our findings indicate that soil and other surfaces containing humic acid exhibit an organic surface photochemistry that produces reductive surface species, which react selectively with nitrogen dioxide. The observed rate of nitrous acid formation could explain the recently observed high daytime concentrations of nitrous acid in the boundary layer, the photolysis of which accounts for up to 60 per cent of the integrated hydroxyl radical source strengths^{3,6–13}. We suggest that this photo-induced nitrous acid production on humic acid could have a potentially significant impact on the chemistry of the lowermost troposphere.

We first studied this photochemically driven conversion of nitrogen dioxide (NO₂) into nitrous acid (HONO) on films of humic acid (HA), representing the complex unsaturated organic materials ubiquitously present in the environment. We exposed these HA films to various NO₂ mixtures in an irradiated flow-tube and detected the NO₂ and HONO gas phase concentrations at its exit. Typically, 1 mg (8 µg cm⁻²) of a given HA was coated onto the inner flow-tube wall. Figure 1a shows the results from a typical experiment, where we irradiated such an organic layer in the 400–700-nm wavelength range

with an irradiance of 162 W m⁻². A flow of synthetic air containing 20 p.p.b. NO₂ at a relative humidity of 20% passed through the reactor (with a residence time of 0.6 s). During the irradiation we observed a substantial loss of NO₂ and a corresponding formation of HONO of similar magnitude, which was a factor of 30 greater than production in the dark. The high conversion yield of about 80% indicates that NO₂ is reduced by photochemically activated electron donors being present in the organic film and not by a catalysed disproportionation reaction¹⁸. No comparable reactivity was observed on clean glass surfaces, in contrast to surfaces containing HAs originating from peat, soil or lignite coal (Supplementary Fig. 1). As a consequence, this photochemically driven conversion is probably common to many surfaces 'rich' in partly oxidized aromatic structures, which facilitate the appearance of photochemically activated electron donors. Because such materials are commonly found on ground but potentially also on airborne surfaces¹⁹ (due to soil abrasion, biomass burning or oxidation of volatile organic compounds), the recognition of their photoreactivity may change our understanding of various processes occurring in the planetary boundary layer.

An experiment under the same conditions as that described above was performed over an extended irradiation time of 13 h. Over the total extent of the irradiation 2×10^{15} molecules of NO₂ per cm² have been reduced on the HA surface (one molecule per 2,500 daltons of HA). This high value indicates that on average each molecule of HA (average molecular weight ~15 kDa) present in the reactor reacted several times with NO₂. However, the reaction is not catalytic because HA is oxidized by reaction with NO₂. This is consistent with the observation that the reactivity of the HA surface decreased by 55%

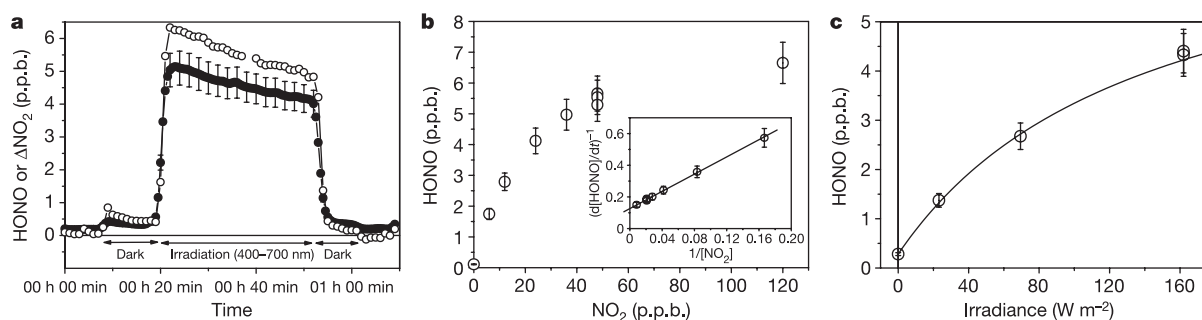


Figure 1 | Conversion of NO₂ → HONO on 1 mg layers of humic acids initiated by visible light (400–700 nm). **a**, The filled circles represent the concentrations of HONO; the empty circles represent the amount of NO₂ removed during an experiment. **b**, The HONO formation as a function of different initial NO₂ concentrations. In the inset the data of the saturation curve is linearized according to a simple photochemical mechanism

(reactions (1)–(3), see Methods section). **c**, The dependence of the HONO formation ([NO₂]₀ = 20 p.p.b.) on the light intensity. The black line is a simple model describing the dependence (see Methods section). The error bars (±2σ) represent the estimated overall accuracy of the chemical analyses.

¹Paul Scherrer Institut, Laboratory of Radio- and Environmental Chemistry, CH-5232 Villigen, Switzerland. ²Bergische Universität Wuppertal, Physikalische Chemie/FB C, D-42097 Wuppertal, Germany. ³Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland. ⁴Laboratoire d'Application de la Chimie à l'Environnement (UCBL-CNRS), F-69622 Villeurbanne, France.

during the 13-h irradiation. HAs are steadily formed by the degradation of biota (about 1 g m^{-2} of soil organic matter is formed per day over the average global continental area²⁰), so the deactivation by NO_2 is small compared to the reformation of HA.

Figure 1b shows the effectiveness of the $\text{NO}_2 \rightarrow \text{HONO}$ conversion at typical tropospheric NO_2 concentrations. The experiments were performed on a single 1 mg HA sample and are corrected for sample deactivation using measured deactivation rates. The HONO yields saturate at high NO_2 concentrations (higher than in the ambient atmosphere) under the reaction conditions described before. An elementary photochemical mechanism—activation of reductive centres (A^{red}) within the organic film by light (reaction (1)), the corresponding deactivation process (reaction (2)), and the reaction of A^{red} with adsorbed NO_2 (reaction (3))—predicts such a saturation (see the Methods section for the mathematical treatment of this reaction system). Compounds 'X' introduced in reactions (1) and (2) indicate that the reactions may involve a photo-induced intra- or intermolecular electron transfer and its back-reaction. Therefore, 'X' should be viewed as oxidants. Clearly, alternative rate-limiting factors, such as a saturation of the adsorption sites by NO_2 , might also explain the observed saturation effect.



The irradiance at the reactor surface in this experiment was about 40% of that corresponding to clear-sky irradiance on a horizontal surface for a solar zenith of 48° in the 300–700-nm range (Supplementary Fig. 2). Thus in the atmosphere the HONO formation can be even more effective.

Figure 1c shows the increase of HONO production with light intensity, which demonstrates the photochemical nature of the reaction. The nonlinearity of the dependency can only be described by reactions (1)–(3), when oxidants 'X' responsible for the deactivation of A^{red} in reaction (2) are photochemically produced transient oxidants (that is, they are formed in reaction (1)). A simple model (see equations (1)–(3) and the Methods section) assuming that the concentrations of oxidants 'X' are proportional to the irradiance is shown in Fig. 1c to match the observations.

The photochemical $\text{NO}_2 \rightarrow \text{HONO}$ conversion occurs over a broad spectral region with approximate overall quantum yields of about 3.9×10^{-6} (300–420 nm), 1.4×10^{-6} (400–700 nm), and 1.5×10^{-6} (mainly 500–700 nm) at light intensities of $144\text{--}162 \text{ W m}^{-2}$, 20 p.p.b. NO_2 and 20% relative humidity (see Supplementary Fig. 2 for lamp spectra). The given values should be taken as a relative measure, because they describe the number of HONO molecules formed per light quantum absorbed by the HA films in the reactor and depend on the experimental parameters (such as thickness of the HA coating or NO_2 concentration). The error in the absolute values is about a factor of two²¹. From the solar spectral distribution, we conclude that the formation of HONO occurs not only in the ultraviolet-A spectral region, where the irradiation can also initiate photodissociation of nitrogen oxides (NO_2 , HONO or $\text{HNO}_3/\text{NO}_3^-$), but is also very effective in the visible region under atmospheric conditions. This is in contrast to the photolysis of nitrate, which has previously been proposed as a daytime source of HONO^{15,22}.

In view of the importance of superoxide as an electron carrier in aquatic photochemistry of dissolved organic matter, it is of interest whether superoxide could be the reductant responsible for the $\text{NO}_2 \rightarrow \text{HONO}$ conversion observed here on solid HA surfaces. Therefore, we performed identical HA irradiations in the absence of NO_2 and measured the yields of H_2O_2 , which is the product of superoxide disproportionation ($2\text{O}_2^- + 2\text{H}^+ \rightarrow \text{H}_2\text{O}_2 + \text{O}_2$). No H_2O_2 was observed under visible light, and only negligible amounts

of H_2O_2 in the ultraviolet-A range (0.05 p.p.b., see Supplementary Fig. 3). Additionally, nitric oxide (NO), which exhibits a reactivity²³ towards superoxide in aqueous solution similar to that of NO_2 , did not react on the HA surfaces under identical conditions (Fig. 1) when using 20 p.p.b. NO instead of NO_2 . The absence of significant amounts of H_2O_2 and the lack of reactivity of NO on irradiated HA surfaces indicate that superoxide is probably not involved in the observed reduction of NO_2 .

HAs are a complex mixture of macromolecular organics. They contain aromatic moieties as visible light absorbers and high contents of phenolic functionalities which can act as electron donors and even show (dark) reactivity towards NO_2 (ref. 24). We have previously shown that films of mixtures of synthetic phenols and an aromatic light absorber (a benzophenone) showed a similar $\text{NO}_2 \rightarrow \text{HONO}$ conversion during ultraviolet-A irradiation, as is observed here for HAs²¹. Such simple mixtures could therefore to some degree be considered as model systems for the photoreactivity of HAs. But we cannot yet assign the exact chemical nature of the reducing intermediates formed on irradiated HAs nor those formed on the synthetic films.

Figure 2a shows the evolution of HONO during irradiation of a

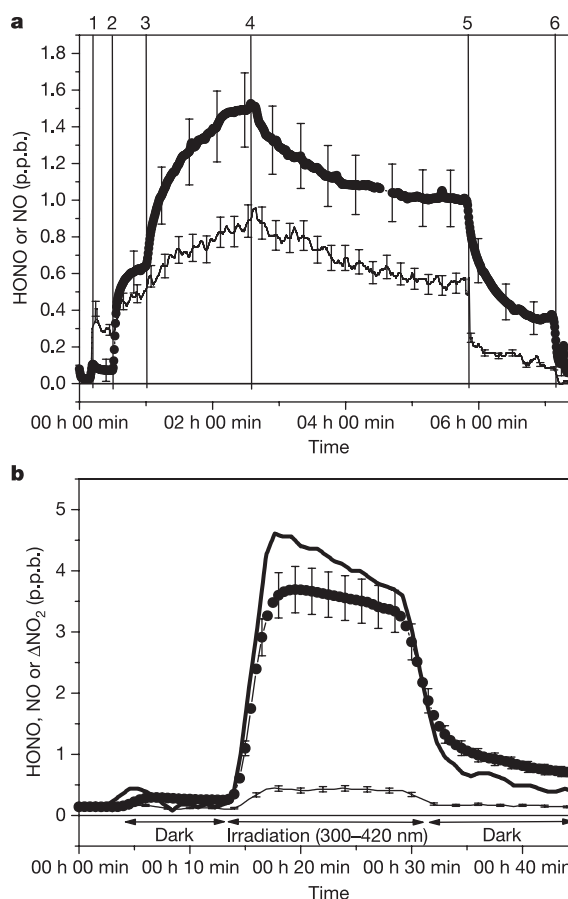


Figure 2 | Conversion of $\text{NO}_2 \rightarrow \text{HONO}$ on irradiated soil in presence of 17 p.p.b. NO_2 and 30% relative humidity. **a**, The formation of HONO (circles) and NO (line) for an ultraviolet-A irradiation (300–420 nm, 70 W m^{-2}) of 6 g of soil. The vertical lines indicate experimental steps 1 to 6: 1: 00 h 06 min, NO_2 added to air flow; 2: 00 h 30 min, soil exposed to NO_2 ; 3: 01 h 00 min, irradiation started; 4: 02 h 40 min, irradiation stopped; 5: 05 h 50 min, NO_2 addition stopped; 6: 07 h 07 min, reactor bypassed. **b**, HONO formation (circles) on $20 \pm 5 \text{ mg}$ artificially acidified soil dust (pH 4.6, H_2SO_4) dispersed on the glass surface of the tubular photo-reactor under ultraviolet-A irradiation (150 W m^{-2}). Thick line, removal of NO_2 ; thin line, formation of NO. The error bars ($\pm 2\sigma$) represent the estimated overall accuracy of the chemical analyses.

natural standard soil (agricultural loamy sand) containing 2.3 wt% organic matter. A 1.5-mm-thick layer of this soil was irradiated (300–420 nm) in a flat-bed flow reactor through a 45 cm × 1 cm glass window by relatively parallel light (70 W m⁻²). While 45 cm² of soil surface were irradiated, a much larger, non-illuminated, internal surface area of at least 5,000 cm² (based on the particle sizes) was actually exposed to NO₂. This meant there was already significant HONO production in the dark owing to NO₂ reduction by HAs (by phenolic moieties²⁴) or NO₂ disproportionation on humid surfaces¹⁸. We suggest that the dark reaction observed under our flow-reactor conditions should not be viewed as representative of soil under environmental conditions, where the transport of NO₂ into the bulk soil is probably much more limited. Upon irradiation the HONO formation was markedly enhanced. The slow response of the HONO concentration to switching the light on and off can be related to its high retention on the large soil surface along the reactor. Under these conditions, NO is also observed as a secondary product of HONO at rates consistent with a known dark reaction of HONO with organic soil constituents^{25,26}. Figure 2b shows HONO formation on a glass surface containing small amounts of soil dust (0.16 mg cm⁻²), demonstrating that surfaces contaminated with traces of soil dust (like roads, buildings, rocks or plants) can also be expected to be photoreactive. Owing to the absence of a large bulk volume of soil in the flow tube, the dark reactions of NO₂ and the retention of HONO on the soil surface are drastically reduced in contrast to the photochemical production.

The light-induced HONO production is 2.5×10^{10} molecules cm⁻² s⁻¹ on soil under ultraviolet-A irradiation (70 W m⁻²) in the presence of 17 p.p.b. NO₂. In the 400–700-nm spectral range we observe HONO photoproduction of 1×10^{10} molecules cm⁻² s⁻¹ for a 75 W m⁻² irradiance, which is low intensity compared to the solar visible irradiance (for example, 400 W m⁻² for 48° solar zenith irradiance). From the experimental results, we estimate the total photochemical HONO production to be 5×10^{10} molecules cm⁻² s⁻¹ for a moderately polluted atmosphere (~20 p.p.b. NO₂) and solar irradiances (300–700 nm) of ~400 W m⁻². For comparison, Staffelbach *et al.*¹⁶ had to introduce an artificial HONO emission of 3.6×10^{10} molecules cm⁻² s⁻¹ to explain the summer daytime HONO concentrations in southern Switzerland with their model. From their measurements they estimate that HONO contributed by more than 30% to the local radical production in air near the ground during the afternoon²⁷. Reports of HONO measurements over a forest¹² and over a rural site⁷ inferred unknown daytime HONO sources of 500 p.p.t. h⁻¹ and 170 p.p.t. h⁻¹, respectively. From the evaluation of the main radical sources—the photolysis of ozone, formaldehyde and HONO or the ozonolysis of alkenes—the authors concluded that HONO photolysis accounted for 33% of the noon-time radical production¹² and for 24% of the 24-h-average radical production⁷. A photochemical HONO formation at the ground surface of 5×10^{10} molecules cm⁻² s⁻¹ is sufficient to establish these HONO source strengths in air columns (assumed to be homogeneously mixed) with heights of 150 m and 430 m, respectively.

Therefore, we conclude that the photochemical HONO formation described here is consistent with recent observed daytime HONO concentrations and can be predicted to have a large contribution (for example, 20–30%) to the OH-radical production of the lowest hundred to a few hundred metres of the atmosphere. A similar impact can also be estimated by a simple OH-radical budget for the lowermost 100 m of the atmosphere: Summer day primary OH production rates of around 10^7 radicals cm⁻³ s⁻¹ are often reported for semi-polluted environments^{7,8,12}. Integrating this OH production over a height of 100 m results in a layer production of 10^{11} radicals cm⁻² s⁻¹. The estimated HONO production of 5×10^{10} molecules cm⁻² s⁻¹ derived in this study could explain half of the observed OH production in this lowermost layer. As the photochemical HONO formation occurs on ground surfaces and HONO photolyses rapidly, this radical source is correspondingly less

important for the formation of OH radicals at higher altitudes in the atmosphere. But this lowest part of the atmosphere is important for the oxidation of biogenic volatile organic compounds, which have a similar short atmospheric lifetime²⁸ as HONO in the daytime atmosphere, and for the formation of secondary air pollutants and aerosols due to the fast radical reactions occurring in this generally most polluted part of the atmosphere.

METHODS

Photoreactors and description of the experiments. The irradiations of HA substrates were performed in 50 cm × 0.8 cm (i.d.) Duran glass tubes installed in an air-cooled lamp-housing holding seven fluorescence lamps (44 cm × 2.6 cm o.d.), in a circular arrangement surrounding the reactor tube. Three types of lamps were used to examine the HONO production under irradiation at different wavelengths (300–420 nm, 400–700 nm and 500–700 nm). The spectral irradiance of the three light sources at the reactor cell surface were measured with a LI-COR 1800 hemispherical, cosine-corrected spectro-radiometer and are shown in Supplementary Fig. 2 and compared to the solar spectral distribution at the Earth's surface and to the absorption spectra of Aldrich-HA. The inner surface of the tubular glass flow reactors (surface = 125 cm², surface to volume ratio = 5 cm⁻¹) was coated with a thin layer of HA. The HA coatings on the reactor wall were produced by drying aliquots of aqueous solutions of the HA (1 mg ml⁻¹, pH = 4.4) dispersed on the reactor walls in a nitrogen stream at room temperature. The experiments with soil were performed in a 45 cm × 1 cm × 1 cm Teflon flow reactor with a glass window (45 cm × 1 cm) at the top. This reactor was illuminated by a single fluorescence lamp (70–75 W m⁻²) mounted in an air-cooled aluminium housing. The flat-bottomed area of the reactor (45 cm²) was completely covered with soil.

Analytical instrumentation. For the measurement of HONO we used a Long Path Absorption Photometer (LOPAP) instrument²⁹ (total accuracy ±10%; ref. 29). NO₂ and NO were detected by means of NO/NO_x-chemiluminescence detectors (CLD, Eco Physics, model CLD 77AM, or Eco Physics model CLD AL 770ppt connected to a photolytic converter PLC 760). The total accuracy of the NO_x measurement is estimated to be ±10%. The NO_x detectors were used in combination with a sodium carbonate denuder tube (50 cm × 0.8 cm) at the inlet of the analysers to remove HONO from the gas stream and therefore eliminate the known interference of HONO in the NO₂ → NO conversion. H₂O₂ was measured with an Aerolaser AL2001CL gas phase monitor.

Materials and reagents. The HA coatings were prepared from the commercially available HA sodium salts from Aldrich. As a possibly photochemically interfering impurity³⁰, the specific lot had an iron content of 0.56%. However, spiking of the HA sample solution by additional 1.2% and 3.6% mass portions of iron(III) did not alter the reactivity of the HA coatings significantly. Standard soil was obtained from the Landwirtschaftliche Untersuchungs- und Forschungsanstalt (LUF), Speyer, Germany. It is a loamy sand (standard soil type Lufa 2.2) collected 15 days before the experiments at 20 cm depth from an agricultural meadow. After drying to 5% residual water content, the soil organic carbon content was $2.29 \pm 0.14\%$ and the soil pH was 5.7 ± 0.3 .

Model calculations used in Fig. 1. The saturation curve in Fig. 1b can be described according to reactions (1)–(3), assuming a steady-state concentration for [A^{red}]_{ss}:

$$\frac{d[\text{HONO}]}{dt} = r(\text{HONO}) = k_3[\text{A}^{\text{red}}]_{\text{ss}}[\text{NO}_2] \quad (4)$$

with

$$[\text{A}^{\text{red}}]_{\text{ss}} = \frac{k_1[\text{HA}]}{k_2[\text{X}] + k_3[\text{NO}_2]} \quad (5)$$

The combination of these two equations results in a HONO formation rate which is first order in NO₂ at low concentrations (that is, $d[\text{HONO}]/dt = k_{\text{eff}}[\text{NO}_2]$), but becomes independent of NO₂ at high NO₂ concentrations (that is, $d[\text{HONO}]/dt = k_{\text{max}}$). A linear regression (equation (6) of the data in a plot of $r(\text{HONO})^{-1}$ versus $[\text{NO}_2]^{-1}$ yields the limiting HONO production $k_{\text{max}} = k_1[\text{HA}]$ at high NO₂ concentrations from the intercept. k_{max} equals the rate of photochemical production of A^{red}. The first-order rate coefficient $k_{\text{eff}} = k_3k_1[\text{HA}]/(k_2[\text{X}])$ for HONO formation at low NO₂ concentrations is derived from the slope. The linearized plot is presented as an inset in Fig. 1b.

$$r(\text{HONO})^{-1} = \frac{k_2[\text{X}]}{k_3k_1[\text{HA}]} \frac{1}{[\text{NO}_2]} + \frac{1}{k_1[\text{HA}]} \quad (6)$$

The available data allow retrieving the maximum HONO production ($k_{\text{max}} = (1.1 \pm 0.2) \times 10^{11}$ molecules s⁻¹ per cm² reactor surface) and the first-order rate coefficient k_{eff} for the HONO formation in the reactor ($k_{\text{eff}} = 0.0048 \pm 0.0007$ s⁻¹ cm⁻²; the given value is normalized to 1 cm² of

reactor surface area). k_{eff} corresponds to a gas kinetic uptake coefficient of $\gamma = 2 \times 10^{-5}$ for the reaction of NO_2 with the HA surface.

In Fig. 1c the parameters k_{max} and k_{eff} derived above are used to model the dependence of HONO formation on the light intensity. Again we use equation (6), but we assume both that the oxidants $[X]$ in reaction (2) are photochemically produced transient oxidants and approximate the concentration of $[X]$ as proportional to the light intensity and that the rate of formation of A^{red} (reaction (1)) is proportional to the light intensity. The model result is shown in Fig. 1c.

Received 22 July 2005; accepted 16 January 2006.

1. Perner, D. & Platt, U. Detection of nitrous acid in the atmosphere by differential optical-absorption. *Geophys. Res. Lett.* **6**, 917–920 (1979).
2. Platt, U., Perner, D., Harris, G. W., Winer, A. M. & Pitts, J. N. Observations of nitrous-acid in an urban atmosphere by differential optical-absorption. *Nature* **285**, 312–314 (1980).
3. Harrison, R. M., Peak, J. D. & Collins, G. M. Tropospheric cycle of nitrous acid. *J. Geophys. Res.* **101**, 14429–14439 (1996).
4. Harris, G. W. *et al.* Observations of nitrous acid in the Los Angeles atmosphere and implications for predictions of ozone-precursor relationships. *Environ. Sci. Technol.* **16**, 414–419 (1982).
5. Lammel, G. & Cape, J. N. Nitrous acid and nitrite in the atmosphere. *Chem. Soc. Rev.* **25**, 361–369 (1996).
6. Alicke, B., Platt, U. & Stutz, J. Impact of nitrous acid photolysis on the total hydroxyl radical budget during the Limitation of Oxidant Production/Pianura Padana Produzione di Ozono study in Milan. *J. Geophys. Res.* **107**, 8196, doi:10.1029/2000JD000075 (2002).
7. Zhou, X. L. *et al.* Summertime nitrous acid chemistry in the atmospheric boundary layer at a rural site in New York State. *J. Geophys. Res.* **107**, 4590, doi:10.1029/2001JD001539 (2002).
8. Alicke, B. *et al.* OH formation by HONO photolysis during the BERLIOZ experiment. *J. Geophys. Res.* **108**, 8247, doi:10.1029/2001JD000579 (2003).
9. Aumont, B., Chervier, F. & Laval, S. Contribution of HONO sources to the $\text{NO}_x/\text{HO}_x/\text{O}_3$ chemistry in the polluted boundary layer. *Atmos. Environ.* **37**, 487–498 (2003).
10. Vogel, B., Vogel, H., Kleffmann, J. & Kurtenbach, R. Measured and simulated vertical profiles of nitrous acid—Part II. Model simulations and indications for a photolytic source. *Atmos. Environ.* **37**, 2957–2966 (2003).
11. Ren, X. R. *et al.* OH and HO_2 chemistry in the urban atmosphere of New York City. *Atmos. Environ.* **37**, 3639–3651 (2003).
12. Kleffmann, J. *et al.* Daytime formation of nitrous acid: A major source of OH radicals in a forest. *Geophys. Res. Lett.* **32**, 05818, doi:10.1029/2005GL022524 (2005).
13. Acker, K. *et al.* Strong daytime production of OH from HNO_2 at a rural mountain site. *Geophys. Res. Lett.* **33**, 02809, doi:10.1029/2005GL024643 (2006).
14. Kleffmann, J. *et al.* Measured and simulated vertical profiles of nitrous acid—Part I: Field measurements. *Atmos. Environ.* **37**, 2949–2955 (2003).
15. Zhou, X. L. *et al.* Nitric acid photolysis on surfaces in low- NO_x environments: Significant atmospheric implications. *Geophys. Res. Lett.* **30**, 2217, doi:10.1029/2003GL018620 (2003).
16. Staffelbach, T., Neftel, A. & Horowitz, L. W. Photochemical oxidant formation over southern Switzerland. 2. Model results. *J. Geophys. Res.* **102**, 23363–23373 (1997).
17. Honrath, R. E. *et al.* Vertical fluxes of NO_x , HONO, and HNO_3 above the snowpack at Summit, Greenland. *Atmos. Environ.* **36**, 2629–2640 (2002).
18. Finlayson-Pitts, B. J., Wingen, L. M., Sumner, A. L., Syomin, D. & Ramazan, K. A. The heterogeneous hydrolysis of NO_2 in laboratory systems and in outdoor and indoor atmospheres: An integrated mechanism. *Phys. Chem. Chem. Phys.* **5**, 223–242 (2003).
19. Krivacsy, Z. *et al.* Study of humic-like substances in fog and interstitial aerosol by size-exclusion chromatography and capillary electrophoresis. *Atmos. Environ.* **34**, 4273–4281 (2000).
20. Janzen, H. H. Carbon cycling in earth systems - a soil science perspective. *Agric. Ecosyst. Environ.* **104**, 399–417 (2004).
21. George, C., Strekowski, R. S., Kleffmann, J., Stemmler, K. & Ammann, M. Photoenhanced uptake of gaseous NO_2 on solid organic compounds: A photochemical source of HONO? *Faraday Discuss.* **130**, 195–210 (2005).
22. Ramazan, K. A., Syomin, D. & Finlayson-Pitts, B. J. The photochemical production of HONO during the heterogeneous hydrolysis of NO_2 . *Phys. Chem. Chem. Phys.* **6**, 3836–3843 (2004).
23. Blough, N. V. in *The Sea surface and global change* (eds Lyss, P. S. & Duce, P. A.) 383–425 (Cambridge University Press, Cambridge, 1997).
24. Ammann, M., Rössler, E., Strekowski, R. & George, C. Uptake of NO_2 on aqueous solutions containing phenoxy type compounds - Implication for HONO formation in the atmosphere. *Phys. Chem. Chem. Phys.* **7**, 2513–2518 (2005).
25. Venterea, R. T. & Rolston, D. E. Mechanisms and kinetics of nitric and nitrous oxide production during nitrification in agricultural soil. *Glob. Change Biol.* **6**, 303–316 (2000).
26. Stevenson, F., Harrison, R. M., Wetselaar, R. & Leeper, R. A. Nitrosation of soil organic matter. 3. Nature of gases produced by reaction of nitrite with lignins, humic substances, and phenolic constituents under neutral and slightly acidic conditions. *Soil Sci. Soc. Am.* **34**, 430–435 (1970).
27. Staffelbach, T. *et al.* Photochemical oxidant formation over southern Switzerland. 1. Results from summer 1994. *J. Geophys. Res.* **102**, 23345–23362 (1997).
28. Atkinson, R. & Arey, J. Gas-phase tropospheric chemistry of biogenic volatile organic compounds: a review. *Atmos. Environ.* **37**, 197–219 (2003).
29. Kleffmann, J., Heland, J., Kurtenbach, R., Lörzer, J. C. & Wiesen, P. A new instrument (LOPAP) for the detection of nitrous acid (HONO). *Environ. Sci. Pollut. Res.* **9**, 48–54 (2002).
30. Zepp, R. G., Faust, B. C. & Hoigne, J. Hydroxyl radical formation in aqueous reactions of iron(II) with hydrogen peroxide—the photo-fenton reaction. *Environ. Sci. Technol.* **26**, 313–319 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Abd El Aal, S. Canonica, M. Birrer, J. Dommen, A. Prévot, L. Urech and I. Alxneit for discussions or technical support. K.S. thanks the Swiss National Science Foundation for support. C.G. acknowledges the grant by Primequal2 for the project SHONO and the CNRS.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.S. (konrad.stemmler@psi.ch).

Cryptic striations in the upper mantle revealed by hafnium isotopes in southeast Indian ridge basalts

D. W. Graham¹, J. Blichert-Toft², C. J. Russo¹, K. H. Rubin³ & F. Albarède²

The Earth's mantle is isotopically heterogeneous on length scales ranging from centimetres to more than 10^4 kilometres^{1,2}. This heterogeneity originates from partial melt extraction and plate tectonic recycling, whereas stirring during mantle convection tends to reduce it. Here we show that mid-ocean ridge basalts from 2,000 km along the southeast Indian ridge (SEIR) display a bimodal hafnium isotopic distribution. This bimodality reveals the presence of ancient compositional striations (streaks) in the Indian Ocean upper mantle. The number density of the streaks is described by a Poisson distribution, with an average thickness of ~ 40 km. Such a distribution is anticipated for a well-stirred upper mantle, in which heterogeneity is continually introduced by plate tectonic recycling, and redistributed by viscous stretching and convective refolding.

The SEIR stretches from the Rodrigues Triple Junction (25.6° S, 70.1° E) to the Macquarie Triple Junction (62° S, 151° E). Between 76°–78° E it crosses the Amsterdam–St Paul (ASP) plateau, a pronounced bathymetric swell associated with relatively hot mantle upwelling beneath the Amsterdam and St Paul islands, while between 120° E and 128° E it crosses the Australian–Antarctic discordance (AAD), a region of deep bathymetry ($>4,000$ m) associated with relatively cold mantle and low melt production³. Notably, over a distance of $\sim 2,500$ km, between 86° E and 120° E, there is a regular eastward decrease in axial depth from 2,300 to 5,000 m, and a morphological transition from axial high to axial valley due to decreasing melt production rate and crustal thickness. This depth gradient occurs at an intermediate and uniform spreading rate (70–75 mm yr⁻¹ full rate) and in the absence of large transform offsets and nearby mantle hotspots. The range in axial depth and ridge morphology is similar to the global range for spreading ridges away from hotspots, making the SEIR a regional-scale analogue of the 50,000-km-long global ocean ridge system. Previous work has established that the He, Pb, Sr and Nd isotope variations along the SEIR are primarily controlled by variation in the depth of melting of isotopically heterogeneous mantle^{4–7}. Also, all SEIR lavas west of the AAD are true 'Indian-type' on the basis of their elevated ²⁰⁸Pb/²⁰⁶Pb ratios⁶.

New Hf isotope results have been obtained for 48 SEIR basalts previously analysed for He–Ne–Ar and Pb–Nd–Sr isotope compositions^{4–7} (see Supplementary Table 1). All samples are fresh mid-ocean ridge basalt (MORB) glasses that were microscopically hand-picked to be free of surface alteration. Between 300 and 600 mg of this glass was digested and the Hf separated using ultrapure reagents and following established techniques⁸. The new results show a +5.5 to +17.8 range in ϵ_{Hf} (defined in Fig. 1). The extreme ϵ_{Hf} values for the data presented here occur on the ASP plateau (+5.5) and in the westernmost AAD (+17.8), and are within the range of values measured previously in those areas^{9–12}. Broadly speaking, Hf and Nd isotopes in our sample suite show the positive correlation typical

of most oceanic basalts¹³. However, in detail the Hf isotope variations are not simply related to axial depth, ridge segmentation, MORB type or Sr–Nd–Pb–He isotopic variations.

The striking attribute of this new data set, previously unobserved in MORBs, is the Hf isotopic bimodality for lavas erupted between 88° E and 110° E (Fig. 2). Over this 2,000-km length of actively spreading ridge the two groupings show a 'gap' of about one epsilon unit ($\epsilon_{\text{Hf}} = +9.5$ to +11.5 and +12.5 to +14.6, respectively). This ϵ_{Hf} gap is significantly larger than the analytical uncertainty (2σ external precision is 0.3 epsilon units). Because our sample suite has a strong spatial resolution, the results suggest the presence of striations (streaks) in the upper mantle beneath the SEIR. The ϵ_{Hf} bimodality is not observed in other geochemical parameters, including Nd isotope composition, indicating that the streaks carry a cryptic memory of ancient chemical fractionation that is now only apparent in the time-integrated Lu/Hf ratio. Because ϵ_{Nd} along this section of the SEIR is not bimodal, Lu–Hf and Sm–Nd fractionation must have been decoupled at some point in the evolutionary history of the underlying upper mantle.

There are several possible origins for the bimodality in Hf isotopes and decoupling of Nd and Hf isotopes, most of which involve mixing with recycled mantle components (Fig. 3). One possibility is a remnant of primordial heterogeneity, perhaps resulting from the presence of a deep magma ocean. Given the efficiency of mantle convection in eradicating such remnants¹⁴, this explanation seems

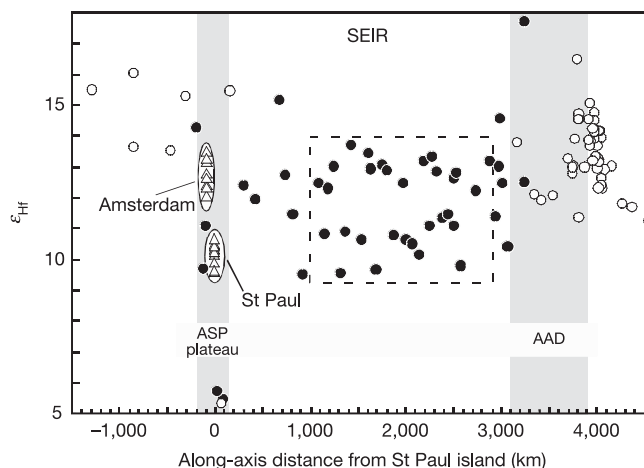


Figure 1 | Along-axis variations in ϵ_{Hf} for basalt glasses from the SEIR. $\epsilon_{\text{Hf}} = ({}^{176}\text{Hf}/{}^{177}\text{Hf}_{\text{measured}}/{}^{176}\text{Hf}/{}^{177}\text{Hf}_{\text{BSE}} - 1) \times 10^4$, where BSE is the bulk silicate Earth reference value of ${}^{176}\text{Hf}/{}^{177}\text{Hf}_{\text{BSE}} = 0.282772$ (ref. 16). New data from this study are shown as solid circles; other data are from refs 9–12. The dashed box outlines the area shown in Fig. 2a.

¹College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. ²Laboratoire des Sciences de la Terre, Ecole Normale Supérieure, 69364 Lyon, France. ³SOEST, University of Hawaii, Honolulu, Hawaii 96822, USA.

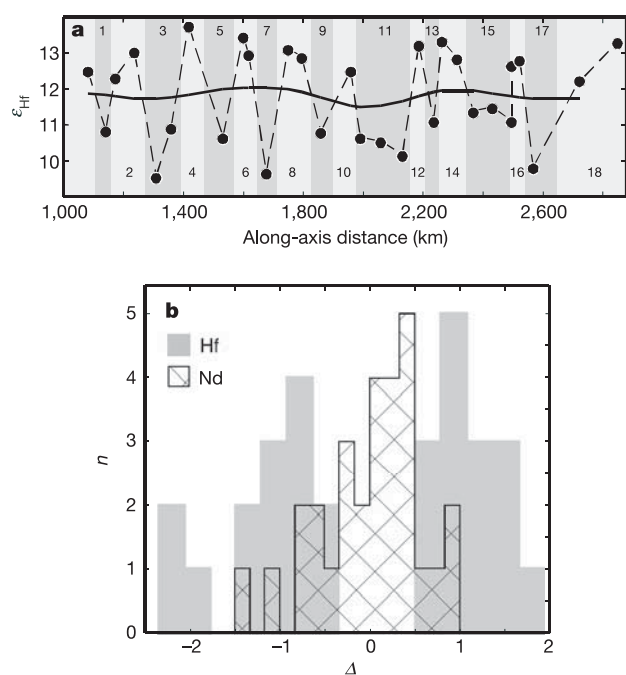


Figure 2 | The Hf isotope bimodality. **a**, Detailed view of the bimodal ϵ_{Hf} region between 88° E and 110° E. The data clearly show two groupings separated by more than one epsilon unit, with high and low values delineated by the alternating stripes numbered 1 to 18 from west to east. The solid curve is a smoothed running mean using a gaussian spatial filter that has a standard deviation of 150 km. **b**, The histogram (n = number of samples) shows the deviations (Δ) from the running mean in **a**. Nd isotopes have been treated similarly to the Hf isotope data and are shown for comparison. Student's t -test indicates that the probability of the Hf isotopes being drawn from a single population is $\ll 0.1\%$.

unlikely, but it cannot be completely ruled out¹⁵. A second possibility is ancient melting that involved a variable amount of garnet restite, such as during komatiite formation¹⁶. The observed north–south ϵ_{Hf} gradient in depleted Atlantic MORB and its relation to distance from the continents can be accounted for in this way^{17,18}.

A third possibility involves mixing of different proportions of tectonically subducted components (altered oceanic crust/lithosphere, pelagic sediment and fluid-modified mantle wedge from subduction zones) with mantle peridotite. Fluid-modified, melt-depleted mantle wedge beneath subduction zones may display ϵ_{Hf} lying above the Hf–Nd mantle array¹⁹. Variability in the mixing proportion of recycled sediment would produce a covariation in ϵ_{Nd} – ϵ_{Hf} , and is therefore unlikely, by itself, to account for the observed Hf isotope bimodality. However, variability in the proportion of pelagic sediment plus hydrothermally altered crust might produce the bimodality, if Hf/Nd ratios in the recycled material range to both higher and lower values than Hf/Nd in the ambient upper mantle (in which case mixing curves in Hf–Nd isotopic space would diverge and could be strongly hyperbolic).

A fourth possibility is a difference in the mineralogical/lithological make-up of the mantle source. For example, clinopyroxene in the residue of mantle melting can develop radiogenic ϵ_{Hf} at relatively invariant ϵ_{Nd} over 10^8 -year timescales, owing to its depletion in high-field-strength elements²⁰. However, there is no evidence for covariation between Hf isotopes in SEIR basalts and potential indicators of modal clinopyroxene abundance or mantle source fertility, such as Sc/Nb or CaO/Al₂O₃ ratios. A fifth possibility is recycling of mid-crustal amphibolite or high-temperature granulite facies rocks in which rutile and iron–titanium oxides control Lu/Hf and the extent of Hf depletion, as evidenced by significant Lu/Hf–Sm/Nd decoupling in crustal xenoliths from southern African kimberlites²¹.

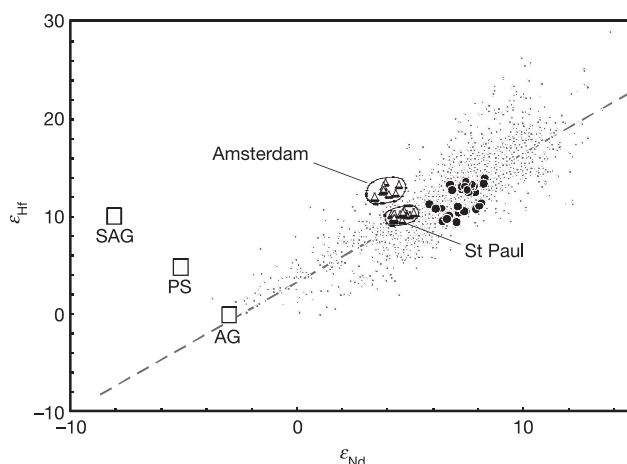


Figure 3 | Global ϵ_{Hf} – ϵ_{Nd} correlation for ~2,100 oceanic basalts. Data are from the literature and from J.B.-T.'s unpublished database. The dashed line is the linear regression for all data ($\epsilon_{\text{Hf}} = 1.3\epsilon_{\text{Nd}} + 3.3$; $R^2 = 0.70$). The bimodal distribution of SEIR basalts is illustrated by the black circles; basalts from the Amsterdam and St Paul islands¹¹, shown by the triangles, display a similar bimodality in ϵ_{Hf} . Examples of potential recycled endmembers that may be involved in mantle mixing are included for comparison (AG = Australian granulite, PS = pelagic sediment, SAG = South African granulite; for example, see ref. 23).

Lastly, garnet in some cratonic peridotites from South African kimberlites shows $\epsilon_{\text{Hf}} = +200$ to $+400$, while ϵ_{Nd} ranges only between 0 and $+15$ (ref. 22). The significant potential for Hf–Nd isotopic decoupling by recycling of continental lithospheric mantle may account for the isotopic variations observed along the southwest Indian ridge²³.

Regardless of their exact origin, the different mantle source regions beneath the SEIR sampled by the MORB Hf isotope bimodality have primarily witnessed a history of coupled Lu/Hf–Sm/Nd fractionation similar to most of the oceanic mantle, because the bimodal populations lie close to the modern ϵ_{Hf} – ϵ_{Nd} array for oceanic basalts and as a group adhere to the overall trend in this diagram (Fig. 3).

Spatially, the streaks defined by the Hf isotope bimodality are well described as a Poisson distribution, in which the number of Hf isotope 'toggles' between the two groupings is proportional to the length of ridge sampled (Fig. 4a). A Poisson distribution describes the total number of independent events that occur within a specified interval for a fixed mean value of the 'arrival rate'; it is characterized by a number of events proportional to the length of the observation span (for example, as exemplified by radioactivity). The number of Hf isotope 'toggles' in SEIR basalts closely follows this prediction over a ridge length of $>2,000$ km, as expected for a Poisson distribution of cryptic streaks in the underlying mantle. Moreover, when the number of striation boundaries follows a Poisson distribution, the intervals between boundaries (that is, the striation thicknesses) follow an exponential distribution, as observed in the current data set (Fig. 4b). The Hf isotope 'toggles' are not as well-discerned outside the 88° E to 110° E section of the SEIR, and seem to be absent where the geodynamic setting is more complex, such as on top of the ASP plateau and within the AAD. Given the sampling density along the SEIR, the mean thickness of the striations in the underlying mantle appears to be ~ 40 km (Fig. 4a inset).

Let us define a geochemical heterogeneity as a point on the boundary between two types of mantle, for example, a mantle residue of partial melting versus recycled crust or lithosphere. A Poisson distribution of geochemical heterogeneities should be anticipated as the natural consequence of a well-stirred upper mantle in which heterogeneities are continually created by tectonic recycling and redistributed by convective stretching and refolding. The mean striation thickness is therefore a useful parameter for quantifying

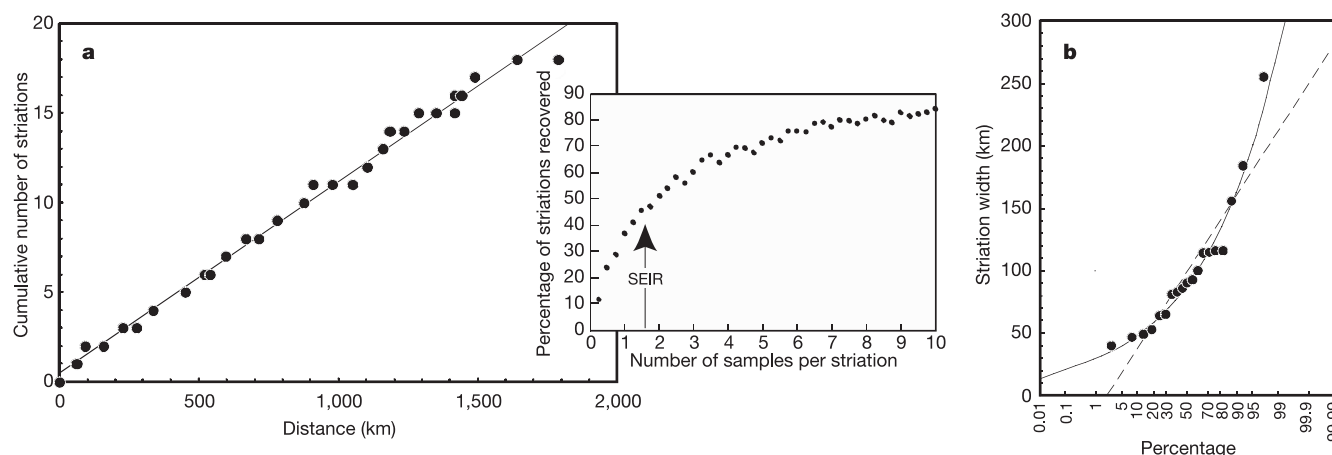


Figure 4 | Poisson character of upper mantle streaks beneath the SEIR. **a**, Cumulative number of Hf isotopic striations versus distance (in kilometres) along the SEIR. The line shown is a linear regression ($y = 0.0107x + 0.487$, $R^2 = 0.992$, $n = 29$; 1σ uncertainties on the slope and intercept are 0.00099 and 0.988, respectively). The mean thickness of the striations, based on the slope of this diagram, would be 93 km (± 9 km, 1σ). The inset illustrates how the inferred mean thickness of the striations is affected by sampling density. A large population with a uniform density for the probability of a transition between two different mantle 'flavours' has been approximated by 1,000 points spread randomly along a line. The distance between two consecutive points represents the thickness of an individual striation. This model has an exponential distribution of striation thicknesses, consistent with the observed relationship between cumulative

number of transitions and distance. As expected, the proportion of striations recovered during sampling increases as the sampling density (that is, number of samples per striation) increases. Because the sampling density along the SEIR has 1.6 samples per striation, about 40% of the striations present in the underlying mantle have been sampled, indicating that the true mean thickness of the striations is ~ 40 km. **b**, Probability diagram of the width (thickness) of the Hf isotopic striations. The x axis shows the percentage of striation widths (normalized to a normal probability distribution) whose value is less than the respective value of the y-axis data point. The solid curve is an exponential fit ($y = 89.71\exp[0.476x]$, $R^2 = 0.971$) and the dashed line is a linear fit ($y = 100.4 + 48.97x$, $R^2 = 0.86$). This good exponential relationship is consistent with a Poisson distribution for the number density of upper mantle striations.

mantle strain rate²⁴. Strain rates associated with mantle convection are of the order of 10^{-15} s^{-1} (ref. 25). After 40 million years (Myr), recycled material would have been stretched and reduced to about 50% of its original thickness via shear strain, and after 300 Myr to about 10%. (Corresponding timescales for stretching by normal strain are shorter, ~ 25 and 80 Myr, respectively; Supplementary Fig. 1.)

The timescale of 300 Myr is similar to that inferred from comparative U–Pb, Sm–Nd and Rb–Sr isotope systematics of enriched and depleted MORB worldwide. This timescale potentially represents a convective cycling time within the mantle, either for oceanic lithosphere and metasomatized mantle subducted beneath island arcs or for continental lithosphere delaminated during rifting, to the eventual resampling by partial melting beneath mid-ocean ridges^{26–28}. Notably, the Pb, Nd, Sr and Hf isotopes in SEIR basalts are correlated with their respective parent/daughter ratios, and each of these correlations also imply 'ages' of 200–400 Myr (ref. 6 and Supplementary Fig. 2). However, these age estimates have large uncertainties, and probably represent minimum values because the slopes of the correlations could also have been affected by varying U/Pb, Sm/Nd, Rb/Sr and Lu/Hf ratios during melting.

The narrow range of ϵ_{Hf} of < 5 units in the bimodal region makes it currently impossible to discriminate between the possible origins of the streaks, because diagnostic geochemical signatures associated with the small extent of Lu/Hf–Sm/Nd decoupling are extremely weak in this sample suite. Nevertheless, the fact that the Amsterdam and St Paul islands display the same Hf isotope bimodality¹¹ as the SEIR MORBs studied here (Fig. 3) suggests a significant role for upwelling plumes in convective dispersion and refolding of such mantle streaks, and is further evidence that regions of the upper mantle far from any active hotspot influence still retain a record of past pollution by mantle plumes²⁹.

The isotopic compositions of mantle-derived materials provide constraints on crust/mantle differentiation and planetary evolution, while their spatial distribution is fundamentally linked to mantle dynamics. The Hf isotope bimodality along the SEIR represents the

first observational evidence that a Poisson distribution of heterogeneous streaks characterizes large sections of the upper mantle. We speculate that this distribution is more easily recognized along the SEIR than along other sections of the mid-ocean ridge system, because the SEIR is uncomplicated by the effects of nearby mantle hotspots, continental land masses, or large fracture zone offsets. Striation thickness distributions in realistic, time-periodic flows of viscous fluids, such as those appropriately describing aspects of the Earth's mantle, are extremely difficult to model numerically given the computational state-of-the-art. However, striation thickness distributions can be accurately predicted from a knowledge of the stretching values³⁰. A Poisson distribution of mantle heterogeneities, and the associated exponential distribution of striation thicknesses, is a fundamental mantle property that should be taken into account in forthcoming models of mantle convection and rheology.

Received 3 August 2005; accepted 12 January 2006.

- Hart, S. R. A large-scale isotope anomaly in the Southern Hemisphere mantle. *Nature* **309**, 753–757 (1984).
- Allègre, C. J. & Turcotte, D. L. Implications of a two-component marble-cake mantle. *Nature* **323**, 123–127 (1986).
- Sempéré, J.-C., Palmer, J., Christie, D., Morgan, J. P. & Shor, A. Australian–Antarctic discordance. *Geology* **19**, 429–432 (1991).
- Graham, D. W., Lupton, J. E., Spera, F. J. & Christie, D. M. Upper mantle dynamics revealed by helium isotope variations along the southeast Indian ridge. *Nature* **409**, 701–703 (2001).
- Nicolaysen, K. E. *The Cretaceous to Recent History of the Southeastern Indian Ocean: The Role of Mantle Plumes Examined by the Geochronology and Sr, Nd, Pb and He Isotopic Geochemistry*. 1–274, PhD dissertation, Mass. Inst. of Technol. (2001).
- Mahoney, J. J. *et al.* Between a hot spot and cold spot: isotopic variation in the Southeast Indian Ridge asthenosphere, 86°–118°E. *J. Petrol.* **43**, 1155–1176 (2002).
- Burnard, P., Graham, D. W. & Farley, K. A. Fractionation of noble gases (He, Ar) during MORB mantle melting: a case study on the Southeast Indian Ridge. *Earth Planet. Sci. Lett.* **227**, 457–472 (2004).
- Blichert-Toft, J., Chauvel, C. & Albarède, F. Separation of Hf and Lu for high-precision isotope analysis of rock samples by magnetic sector-multiple collector ICP-MS. *Contrib. Mineral. Petrol.* **127**, 248–260 (1997).

9. Chauvel, C. & Blichert-Toft, J. A hafnium and trace element perspective on melting of the depleted mantle. *Earth Planet. Sci. Lett.* **190**, 137–151 (2001).
10. Kempton, P. L., Pearce, J. A., Barry, T. L., Langmuir, C. H. & Christie, D. M. Sr-Nd-Pb-Hf isotope results from ODP Leg 187: evidence for mantle dynamics of the Australian–Antarctic discordance and origin of the Indian MORB source. *Geochem. Geophys. Geosyst.* **3**, doi:10.1029/2002GC000320 (2002).
11. Doucet, S., Weis, D., Scoates, J. S., Debaille, V. & Giret, A. Geochemical and Hf-Pb-Sr-Nd isotopic constraints on the origin of the Amsterdam–St. Paul (Indian Ocean) hotspot basalts. *Earth Planet. Sci. Lett.* **28**, 179–195 (2004).
12. Hanan, B., Blichert-Toft, J., Pyle, D. G. & Christie, D. M. Contrasting origins of the upper mantle revealed by hafnium and lead isotopes from the Southeast Indian Ridge. *Nature* **432**, 91–94 (2004).
13. Salters, V. J. M. & White, W. M. Hf isotope constraints on mantle evolution. *Chem. Geol.* **145**, 447–460 (1998).
14. Blichert-Toft, J. & Albarède, F. Short-lived chemical heterogeneities in the Archaean mantle with implications for mantle convection. *Science* **263**, 1593–1596 (1994).
15. Albarède, F. in *Structure, Composition and Evolution of the Earth's Mantle* (eds van der Hilst, R. D., Bass, J., Matas, J. & Trampert, J.) 27–46 (Geophys. Monogr. 160, American Geophysical Union, Washington DC, 2005).
16. Blichert-Toft, J., Arndt, N. & Gruau, G. Hf isotope measurements on Barberton komatiites: effects of incomplete sample dissolution and importance for primary and secondary magmatic signatures. *Chem. Geol.* **207**, 261–275 (2004).
17. Andres, M., Blichert-Toft, J. & Schilling, J.-G. Nature of the depleted upper mantle beneath the Atlantic: evidence from Hf isotopes in normal mid-ocean ridge basalts from 79°N to 55°S. *Earth Planet. Sci. Lett.* **225**, 89–103 (2004).
18. Blichert-Toft, J. *et al.* Geochemical segmentation of the Mid-Atlantic Ridge north of Iceland and ridge-hotspot interaction in the North Atlantic. *Geochem. Geophys. Geosyst.* **6**, doi:10.1029/2004GC000788 (2005).
19. Woodhead, J. D., Hergt, J. M., Davidson, J. P. & Eggins, S. M. Hafnium isotope evidence for 'conservative' element mobility during subduction zone processes. *Earth Planet. Sci. Lett.* **192**, 331–346 (2001).
20. Salters, V. J. M. & Zindler, A. Extreme $^{176}\text{Hf}/^{177}\text{Hf}$ in the sub-oceanic mantle. *Earth Planet. Sci. Lett.* **129**, 13–30 (1995).
21. Schmitz, M. D., Vervoort, J. D., Bowring, S. A. & Patchett, P. J. Decoupling of the Lu–Hf and Sm–Nd isotope systems during evolution of granulitic lower crust beneath southern Africa. *Geology* **32**, 405–408 (2004).
22. Bedini, R. M., Blichert-Toft, J., Boyet, M. & Albarède, F. Isotopic constraints on the cooling of the continental lithosphere. *Earth Planet. Sci. Lett.* **223**, 99–111 (2004).
23. Janney, P. E., le Roex, A. P. & Carlson, R. W. Hafnium isotope and trace element constraints on the nature of mantle heterogeneity beneath the central Southwest Indian Ridge (13°E to 47°E). *J. Petrol.* **46**, 2427–2464 (2005).
24. Olson, P. L., Yuen, D. A. & Balsiger, D. S. Mixing of passive heterogeneities by mantle convection. *J. Geophys. Res.* **89**, 425–436 (1984).
25. Schubert, G., Turcotte, D. L. & Olson, P. *Mantle Convection in the Earth and Planets* 1–940 (Cambridge Univ. Press, Cambridge, Massachusetts, 2001).
26. Allègre, C. J. & Lewin, E. Isotopic systems and stirring times of the Earth's mantle. *Earth Planet. Sci. Lett.* **136**, 629–646 (1995).
27. Albarède, F. Radiogenic ingrowth in systems with multiple reservoirs: applications to the differentiation of the mantle-crust system. *Earth Planet. Sci. Lett.* **189**, 59–73 (2001).
28. Donnelly, K. E., Goldstein, S. L., Langmuir, C. H. & Spiegelman, M. Origin of enriched ocean ridge basalts and implications for mantle dynamics. *Earth Planet. Sci. Lett.* **226**, 347–366 (2004).
29. Hanan, B. B., Kingsley, R. H. & Schilling, J.-G. Pb isotope evidence in the South Atlantic for migrating ridge–hotspot interactions. *Nature* **322**, 137–144 (1986).
30. Muzzio, F. J., Alvarez, M. M., Cerbelli, S., Giona, M. & Adrover, A. The intermaterial area density generated by time- and spatially periodic 2D chaotic flows. *Chem. Eng. Sci.* **35**, 1497–1508 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements D.W.G. thanks A. Bouvier, A. Agranier, P. Beck, F. Moynier, K. Koga and E. Koga for their help and hospitality during two extended visits to Lyon. We thank P. Telouk for his help with the P54 mass spectrometer. D. Christie, B. Hanan, K. Johnson, J. Mahoney and D. Pyle provided help on many aspects of SEIR geochemistry. N. Pisiás provided insight into some of the nuances of statistical distributions. D.W.G., C.J.R. and K.H.R. were supported by the Marine Geology division of the NSF, and J.B.-T. and F.A. by the French Institut National des Sciences de l'Univers.

Author Contributions D.W.G. and J.B.-T. performed the Hf isotope measurements. All authors contributed to data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.W.G. (dgraham@coas.oregonstate.edu).

Stratified prokaryote network in the oxic–anoxic transition of a deep-sea halocline

Daniele Daffonchio¹, Sara Borin¹, Tullio Brusa¹, Lorenzo Brusetti¹, Paul W. J. J. van der Wielen², Henk Bolhuis², Michail M. Yakimov³, Giuseppe D'Auria³, Laura Giuliano³, Danielle Marty⁴, Christian Tamburini⁴, Terry J. McGenity⁵, John E. Hallsworth⁵, Andrea M. Sass⁵, Kenneth N. Timmis^{5,6}, Anastasios Tselepidis⁷, Gert J. de Lange⁸, Andreas Hübner⁸, John Thomson⁹, Soterios P. Varnavas¹⁰, Francesco Gasparoni¹¹, Hans W. Gerber¹², Elisa Malinverno¹³, Cesare Corselli¹³ & Biodeep Scientific Party*

The chemical composition of the Bannock basin has been studied in some detail^{1,2}. We recently showed that unusual microbial populations, including a new division of Archaea (MSBL1)³, inhabit the NaCl-rich hypersaline brine. High salinities tend to reduce biodiversity⁴, but when brines come into contact with fresher water the natural haloclines formed frequently contain gradients of other chemicals, including permutations of electron donors and acceptors, that may enhance microbial diversity, activity and biogeochemical cycling^{5,6}. Here we report a 2.5-m-thick chemocline with a steep NaCl gradient at 3.3 km within the water column between Bannock anoxic hypersaline brine⁷ and overlying sea water. The chemocline supports some of the most biomass-rich and active microbial communities in the deep sea, dominated by Bacteria rather than Archaea, and including four major new divisions of Bacteria. Significantly higher metabolic activities were measured in the chemocline than in the overlying sea water and underlying brine; functional analyses indicate that a range of biological processes is likely to occur in the chemocline. Many prokaryotic taxa, including the phylogenetically new groups, were confined to defined salinities, and collectively formed a diverse, sharply stratified, deep-sea ecosystem with sufficient biomass to potentially contribute to organic geological deposits.

High-precision sampling was conducted during cruises of the research vessel *Urania* equipped with the Modus–Scipack system (<http://www.geo.unimib.it/BioDeep/Project.html>; Fig. 1a). The vehicle Modus, connected by cable to the research vessel, held a second instrument, the Scipack, with a 10-m data transmission cable. The Scipack, consisting of a Rosette sampler equipped with a CTD (conductivity–temperature–depth probe) and a series of Niskin bottles, was connected to the Modus through the Sciskid, a module equipped with a pressure sensor for recording the pressure at which the Niskin bottles were closed (Fig. 1c). A camera on the Modus could provide an image of the Scipack entering the brine lake (Fig. 1b, and Supplementary Fig. S1). Immediately after sampling, the Modus–Scipack was raised, the Niskin bottles were retrieved and their contents were carefully fractionated on board ship by slowly recovering 0.5-litre, 1-litre or 2-litre fractions from the bottom tap. These were then immediately analysed for salinity (Fig. 1d). The reconstructed interface salinity profile was strongly positively

correlated ($r = 0.98$, $P < 0.001$) with the CTD conductivity profile recorded in independent non-sampling casts (Fig. 2d), indicating that little or no mixing had occurred.

The interface halocline was about 2.5 m deep, in agreement with previous estimates that employed alternative sampling strategies¹. Although biomass values fluctuated along the halocline, there were significantly greater numbers of microbial cells in the interface (about 10^6 cells ml⁻¹) than in either the deep sea water or the underlying hypersaline brine, both of which had about 10^4 cells ml⁻¹ (Fig. 2a). Comparison of archaeal and total prokaryotic 16S ribosomal RNA gene abundance revealed that Bacteria and Archaea were present at equivalent levels in the oxic seawater above the hypersaline brine, as reported previously for the deep sea⁸. In contrast, the interface samples contained 20.5 and 1.8 pg ml⁻¹ bacterial and archaeal 16S rRNA gene, respectively (Fig. 2b), contrary to the notion that extreme environments are dominated by Archaea but consistent with the low level of Archaea-mediated methanogenesis measured in the upper 2 m of the halocline (Fig. 2c). The high concentration of bacteria in the interface indicates that it acts as a barrier to particulates that sink through the marine water column, leading to an accumulation of organic carbon, nutrients and reactive surfaces that support microbial growth⁹. In addition, methane originating in the hypersaline brine will rise and enter the interface, where it can be oxidized under more energetically favourable conditions.

Levels of ATP, an important indicator of metabolic activity¹⁰, were significantly higher in the interface than in the overlying sea water and the brine lake (Fig. 2b), as were extracellular aminopeptidase and alkaline phosphatase activities (Supplementary Table S1). At certain depths of the interface, sulphate reduction rates (SRRs) showed appreciably higher values than those of the deep oxic sea water¹¹, but the highest SRRs were measured in the underlying anoxic brine (Fig. 2c).

Down the halocline, nitrate concentration decreased from 6.6 to 0.3 μ M, ammonium increased from 5 to 3,450 μ M (refs 2, 12), sulphate, the most abundant electron acceptor along the halocline, increased from 31 to 84 mM, manganese increased from 0.4 to 8.3 μ M and redox potential decreased from 210 mV to less than zero, indicating a total depletion of oxygen in the lower part of the chemocline (Fig. 2e–g). Electron acceptors and donors potentially

¹CoNISMa, UIR Università degli Studi di Milano, DiSTAM, 20133 Milan, Italy. ²Department of Microbial Ecology, CEES, University of Groningen, 9751 NN Haren, The Netherlands.

³Istituto per l'Ambiente Marino Costiero, CNR, 98122 Messina, Italy. ⁴LMGEM, UMR 6117 CNRS Université de la Méditerranée, 13288 Marseille, France. ⁵Department of Biological Sciences, University of Essex, Colchester, Essex CO4 3SQ, UK. ⁶Division of Microbiology, GBF, 38122 Braunschweig, Germany. ⁷Institute of Marine Biology of Crete, 71003 Heraklion, Greece. ⁸Faculty of Geosciences, Geochemistry, Utrecht University, 3584 CD Utrecht, The Netherlands. ⁹Southampton Oceanography Centre, Southampton SO14 3EE, UK. ¹⁰Department of Geology, University of Patras, 26100 Patras, Greece. ¹¹Tecnomare S.p.A., ENI Group, 30124 Venice, Italy. ¹²Technische Fachhochschule Berlin, University of Applied Science, 13353 Berlin, Germany. ¹³CoNISMa, UIR Università degli Studi di Milano Bicocca, DGSG, 20126 Milan, Italy.

*A list of participants and affiliations appear at the end of the paper.

available for metabolic processes in the interface are thus variously distributed along the gradient. Whereas Na^+ showed chemically conservative behaviour along the halocline, dissolved manganese, nitrate, ammonium and sulphate concentrations exhibited nonlinear slopes, demonstrating non-conservative behaviour presumably reflecting biologically mediated redox cycling¹³ (Supplementary Fig. S2).

Detailed vertical profiles of biological parameters showed two maxima in diversity at about 8‰ and 15–22‰ salinity, which corresponded to maxima in biomass, determined by total cell counts and prokaryote 16S rRNA gene abundance analysis ($r = 0.93$, $P < 0.001$; Fig. 2a), and in ATP concentration ($r = 0.74$, $P < 0.01$ and $r = 0.64$, $P < 0.05$; Fig. 2b). The decreases in these biological parameters in the 10–15‰ salinity range, and the low SRR in this region, can be explained by a depletion of nutrients and organic substrates from above and below, and/or a localized shortage of energetically favourable redox couplings (for example nitrate and oxygen are depleted at this point; Fig. 2e, f).

NaCl is well known for generating high osmotic potentials, reducing the mole fraction of water, and its kosmotropic (order-generating) activity stabilizes cellular macromolecules and stiffens membranes (Supplementary Fig. S3): such parameters influence the environmental windows of growth for any microbial species. In contrast with other hypersaline environments, such as coastal solar salt pans⁴, bacterial diversity in the Bannock interface was higher than that found in the adjacent low-salt habitat (the overlying deep seawater), as shown by Shannon–Weaver indices derived from

amplified ribosomal intergenic spacer analysis (ARISA) of Bacteria at 18 different salinities (Fig. 2a). To identify the Bacteria present at different locations in the halocline, clone libraries of RT–PCR amplicons of 16S rRNA (complementary DNA) and of PCR amplicons of 16S rRNA genes were generated; 539 clones, including 179 cDNA clones representing the more metabolically active microbial fraction, were sequenced. This analysis confirmed the high bacterial phylogenetic diversity suggested by the Shannon–Weaver indices and identified four new candidate divisions designated Mediterranean Sea Brine Lake groups 3–6 (MSBL3–6) and a dominant candidate division, newly named MSBL2 (Supplementary Fig. S4). Candidate division MSBL2 is phylogenetically related to the candidate division SB1, previously identified in the interface of the Shaban Deep in the Red Sea¹⁴. MSBL2-related sequences were also detected in the interfaces of other hypersaline basins, but were absent from the underlying brines or the overlying sea water³, which indicates that bacteria belonging to the MSBL2 division are specifically adapted to the seawater–hypersaline brine interfaces.

The deeper part of the halocline (22.2–25.0‰ salinity) was inhabited by MSBL2, by Sphingobacteriaceae and by candidate divisions KB1 (refs 14, 15) and MSBL3–6 (Fig. 3a). The candidate division KB1 includes species related to Thermotogales that have been found in brine lakes and their sediments in the Red Sea^{14,15}, and in other anaerobic hypersaline sediments¹⁶. Despite being less abundant than those of MSBL2, bacteria of MSBL3–6 candidate divisions were found only in the deeper part of the halocline, indicating that they are adapted to high salinity. Archaeal clones

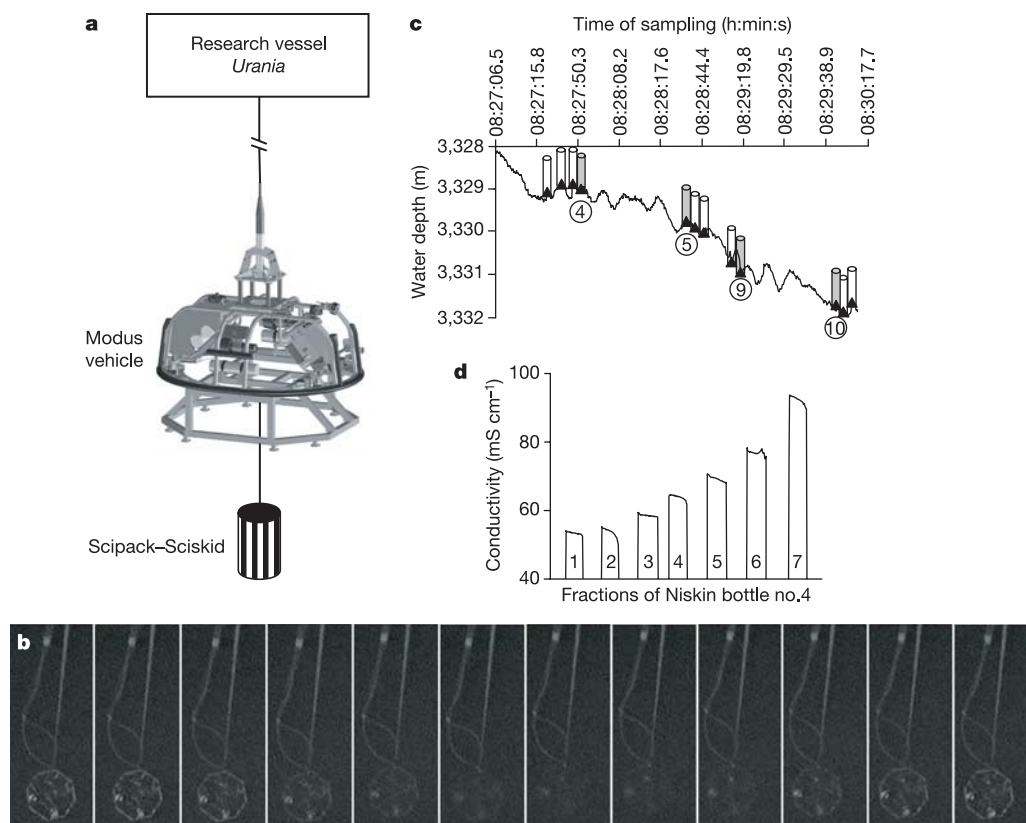


Figure 1 | Procedures for sample recovery from the seawater–Bannock brine interface. **a**, Schematic drawing of the Modus–Scipack system illustrating the Modus vehicle holding the Scipack–Sciskid unit consisting of a Rosette sampler equipped with a CTD and a series of Niskin bottles (Scipack), and a pressure sensor (Sciskid) for recording the pressure at which the Niskin bottles were closed. **b**, Photographs of the Scipack Rosette sampler at 0.5-s intervals while going into and out of the seawater–Bannock brine interface obtained during a video survey from the Modus vehicle (from left to right). The disappearance of the Scipack is caused by the

transmittance change resulting from the high salinity of the brine. **c**, This sampling system yielded the exact time and depth at which each Niskin bottle (represented by cylinders over the curve) was closed. In this way it was possible to determine at which point of the conductivity curve during the cast the bottles were closed. The circled numbers below selected bottles identify those Niskin bottles (shaded) that were subsampled on board ship for the characterization of the gradient. **d**, Example of conductivity measurements of the different water fractions sub-sampled from Niskin bottle no. 4 reported in **c**, using the conductivity probe from the Scipack.

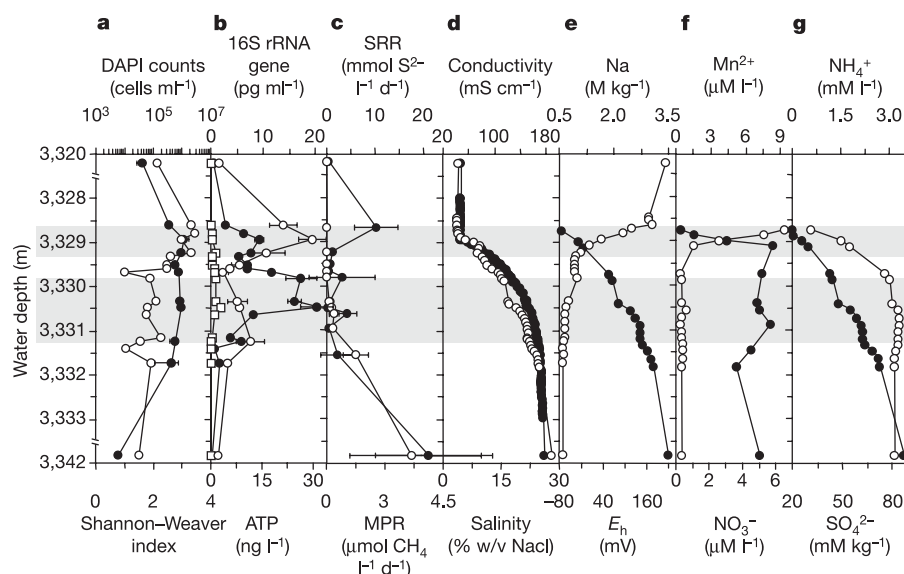


Figure 2 | Microbiological and geochemical profiling of the seawater-Bannock brine interface. **a**, Filled circles, DAPI microbial counts; open circles, Shannon-Weaver indices calculated from ARISA data. **b**, Filled circles, prokaryote 16S rRNA gene abundance; open squares, Archaea 16S rRNA gene abundance; open circles, ATP. **c**, Filled circles, SRR; open circles, methane production rate (MPR). **d**, Filled circles, conductivity; open circles, salinity. **e**, Filled circles, sodium; open circles, redox potential (E_h). **f**, Filled circles, dissolved manganese; open circles, nitrate. **g**, Filled circles,

ammonium; open circles, sulphate. Shading indicates zones with relatively high biological activity. Conductivity data in **c** were obtained during a non-sampling cast by continuous measurements with a conductivity probe mounted on the Scipack system. All the data were from station AB27SCI except values for deep sea water and Bannock brine that were from station AB29SCI. Error bars indicate standard deviations. For sodium, dissolved manganese, nitrate, ammonium and sulphate, standard deviations were 2–5% of each value.

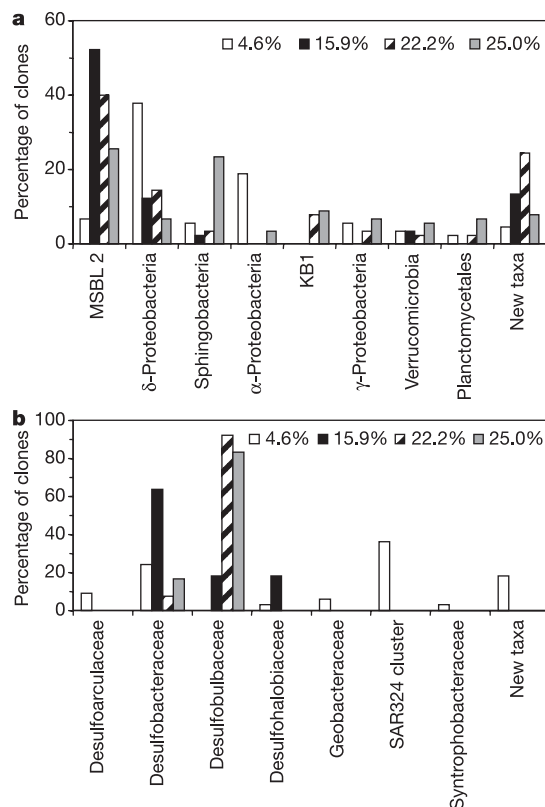


Figure 3 | Distribution of prokaryotic diversity at different salinities along the seawater-Bannock brine interface. **a**, Percentage distribution of 16S rRNA gene clones belonging to different bacterial phylogenetic groups. New divisions MSBL3–6 are included under 'new taxa'. **b**, Percentage distribution of different families of the δ -Proteobacteria.

related to members of candidate divisions MSBL1, previously found to dominate in the brine³, and ANME-1, associated with anaerobic oxidation of methane¹⁷, were found in cDNA libraries generated from samples from the deeper part of the halocline, namely the anoxic zone where methane and sulphate coexist. Sulphate-reducing δ -Proteobacteria were abundant, comprising 6–15% of the clones in the libraries from all anoxic levels of the interface. Desulfobacteriaceae and Desulfobulbaceae were found mainly at lower and higher salinities, respectively (Fig. 3b), indicating specific adaptation of these two groups to the different prevailing environmental factors of salinity, redox potential, oxygen^{18,19}, sulphate concentration and availability of organic substrates.

Eighty-four bacterial isolates were cultured from the Bannock interface; on the basis of partial 16S rRNA sequences they were assigned to the Firmicutes, Bacteroidetes, α -Proteobacteria, γ -Proteobacteria and ϵ -Proteobacteria (Supplementary Table S2). The isolates belonged to divisions widely represented in the rRNA libraries; however, they did not correspond to any specific sequence found in the libraries (Supplementary Fig. S4). Most isolates were moderately halophilic facultative anaerobes, able to grow aerobically or by fermentation or denitrification, and therefore well adapted to the seawater-brine interface. In particular, a *Halothiobacillus* species was able to aerobically oxidize thiosulphate, with CO_2 as sole carbon source, over a NaCl range of 0.5–23%. In addition, two obligately anaerobic, fermentative, halophilic isolates had 99.8% 16S rRNA sequence similarity to *Halanaerobium* sp. KT-2/3-3, which was isolated from a similar seawater-brine interface in the Kebrit deep²⁰.

Two obligately anaerobic strains, one from the Bacteroidetes and one from the ϵ -Proteobacteria from the same enrichment, had less than 92% 16S rRNA sequence similarity to known organisms. Both were moderate halophiles, growing between 2% and 12% NaCl and between 0.1% and 9.0% NaCl, respectively. The Bacteroidetes isolate fermented a variety of sugars and biopolymers as sole sources of carbon and energy, whereas the ϵ -Proteobacteria member fermented a variety of organic acids, and used nitrate, nitrite, thiosulphate, trimethylamine *N*-oxide, dimethylsulphoxide and elemental sulphur as terminal electron acceptors with formate and acetate as electron

donors. It can therefore be envisaged that these organisms, which are highly suited to life in Bannock interface, would benefit mutually from the degradation of organic polymers, which are likely to accumulate at the interface⁹, coupled to the respiration of sulphur compounds or nitrate.

Sulphate reducers, despite being abundantly represented in clone libraries, escaped cultivation. Nevertheless, analysis of our isolates confirmed that a range of biological processes—such as fermentation and the oxidation of reduced sulphur species and organic matter coupled to a variety of terminal electron acceptors—are likely to occur in the interface.

In conclusion, the Bannock brine lake–seawater interface is revealed here to be a deep-sea microbial oasis, consisting of a biomass-dense, metabolically active microbial ecosystem that receives methane from the hypersaline brine below and organic particulates sedimenting through the water column from above, recycles this organic matter and interconverts diverse ionic species, which are used and reused as electron acceptors and donors. Although functionally analogous to the deep-sea floor, which also collects and recycles sinking particulates, the interfaces of deep hypersaline basin lakes differ significantly in the stress imposed by their steep halocline, and consequently have been selected for stratified microbial communities containing new organisms, including divisions of bacteria not previously described. Recent work detailing the retrieval of novel enzymes from the analogous Urania basin interface²¹ would suggest that this new phylogenetic diversity reflects new functional activity. Such biomass-rich environments are stable over hundreds of years, constituting localized sources of organic matter that may contribute to the geological record, for example in the form of sapropels^{22,23}.

METHODS

Sampling of seawater–brine interface in Bannock basin. Sampling of the Bannock basin was conducted from the research vessel *Urania* at location 34° 21.640' N, 20° 02.260' E in 2001, location 34° 17.949' N, 20° 00.985' E (station BD29CT) in 2002, and at locations 34° 17.488' N, 20° 00.692' E (AB27SCI) and 34° 17.397' N, 20° 00.709' E (AB29SCI) in 2003. The halocline water fractions, sampled with the Modus–Scipack system, were analysed as reported below for microbial activity, microbial abundance and diversity, physicochemical properties, and isolation of microbial strains.

Activity measurements. Aminopeptidase and phosphatase activities were determined with the fluorogenic substrates L-leucine-7-amino-4-methylcoumarin and 4-methylumbelliferyl phosphate with the use of the multi-concentration kinetic method at concentrations ranging from 0.05 to 10 μ M (ref. 24). Methane production rates were determined by measuring the production of methane with a gas chromatograph equipped with a flame ionization detector³. Sulphate reduction rates were determined by measuring the [³⁵S]sulphide production from radiolabelled sulphate (1–2 μ Ci [³⁵S]sulphate) with standard methods. ATP was measured on triplicate 10-ml samples filtered through 0.22- μ m pore-size filters. ATP was extracted and measured directly on the filter with the luciferin–luciferase-based biomass test kit (Promicol), with a luminometer, following the instructions of the manufacturer. Relative luminescence units were converted to ATP concentrations with the use of a standard ATP curve.

DNA/RNA-based analysis and direct cell count. Samples were filtered onto sterile 0.22- μ m pore-size filters that were stored at –20 °C in 2 ml of sterile lysis buffer (EDTA 40 mM, Tris–HCl 50 mM pH 5.8, sucrose 0.75 M). DNA was extracted from each filter as described previously³. Extracted DNA was quantified by comparative agarose-gel electrophoresis with known quantities of lambda phage DNA. 4,6-Diamidino-2-phenylindole (DAPI) staining was performed as described previously³. Real-time PCR experiments for the quantification of total Prokaryotes and Archaea were performed by Taqman assays as described elsewhere²⁵.

From each DNA fraction along the salinity gradient, ARISAs were performed in accordance with an established protocol²⁶. As a parameter for the structural diversity of a bacterial community a Shannon–Weaver index²⁷ (H) was calculated for representative ARISA profiles by using the function $H = -\sum (n_i/N) \ln(n_i/N)$, where n_i is the height of a peak and N is the sum of all peak heights in an ARISA profile. The construction and sequencing of 16S rRNA gene libraries from fractions with different salinities were performed with

a previously described procedure³. Sequences were aligned with the use of ARB software²⁸, and operational taxonomic unit (OTU) distribution at different salinities was calculated. Good coverage of the dominant OTU population was confirmed with rarefaction analysis of the clone libraries. Sequences of 16S rRNA genes showing more than 97% homology were considered to belong to the same OTU. For RNA extraction and reverse transcription the RNA/DNA Mini Isolation kit (Qiagen) and Superscript II reverse transcriptase (Invitrogen) were used.

Geochemical analyses and salinity-related measurements. The redox potential in each halocline fraction was measured on board ship with a portable E_h meter²⁹. Major elements in brines were analysed by inductively coupled plasma atomic emission spectrometry at Geosciences Utrecht after dilution to 3.5% salinity; nitrate and ammonia were analysed by AutoAnalyser at NIOZ (Nederlands Instituut voor Onderzoek der Zee), Texel; Mn was measured at RIVM (Rijks Instituut voor Volksgezondheid en Milieu), Bilthoven, with the use of high-resolution inductively coupled plasma mass spectrometry.

Bacterial isolation and characterization. A collection of 84 isolates was obtained by inoculating anaerobic samples of the whole interface into a wide variety of media (Supplementary Information). Isolates were tentatively identified by sequencing the 16S rRNA gene, and physiological features tested included growth at different salinities and with different carbon and energy sources and terminal electron acceptors.

Received 18 October; accepted 9 November 2005.

- De Lange, G. J., Catalano, G., Klinkhammer, G. P. & Luther, G. W. The interface between oxic seawater and the anoxic Bannock brine: its sharpness and the consequences for the redox-related cycling of Mn and Ba. *Mar. Chem.* **31**, 205–217 (1990).
- De Lange, G. J. *et al.* Composition of anoxic hypersaline brines in the Tyro and Bannock basins, eastern Mediterranean. *Mar. Chem.* **31**, 63–88 (1990).
- Van der Wielen, P. W. J. J. *et al.* The enigma of prokaryotic life in deep hypersaline anoxic basins. *Science* **307**, 121–123 (2005).
- Benlloch, S. *et al.* Prokaryotic genetic diversity throughout the salinity gradient of a coastal saltern. *Environ. Microbiol.* **4**, 349–360 (2002).
- D'Hondt, S. *et al.* Distribution of microbial activities in deep seafloor sediments. *Science* **306**, 2216–2221 (2004).
- Parkes, R. J. *et al.* Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436**, 390–394 (2005).
- Scientific Staff of Cruise Bannock 1984–12. Gypsum precipitation from cold brines in an anoxic basin in the eastern Mediterranean. *Nature* **314**, 152–154 (1985).
- Karner, M. B., De Long, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–511 (2001).
- Henneke, E. & De Lange, G. J. The distribution of DOC and POC in the water column and brines of the Tyro and Bannock basins. *Mar. Chem.* **31**, 113–122 (1990).
- LaRock, P. A., Lauer, R. D., Schwarz, J. R., Watanabe, K. K. & Wiesenburg, D. A. Microbial biomass and activity distribution in an anoxic hypersaline basin. *Appl. Environ. Microbiol.* **37**, 466–470 (1979).
- Luther, G. W., Catalano, G., De Lange, G. J. & Woittiez, J. R. W. Reduced sulphur in the hypersaline anoxic basins of the Mediterranean Sea. *Mar. Chem.* **31**, 137–152 (1990).
- Henneke, E. & De Lange, G. J. Sulphur speciation in anoxic hypersaline sediments from the eastern Mediterranean Sea. *Geochim. Cosmochim. Acta* **61**, 307–321 (1997).
- Van Cappellen, P. *et al.* Biogeochemical cycles of manganese and iron at the oxic–anoxic transition of a stratified marine basin (Orca Basin, Gulf of Mexico). *Environ. Sci. Technol.* **32**, 2931–2939 (1998).
- Eder, W., Schmidt, M., Koch, M., Garbe-Schönberg, D. & Hubner, R. Prokaryotic phylogenetic diversity and corresponding geochemical data of the brine–seawater interface of the Shaban Deep, Red Sea. *Environ. Microbiol.* **4**, 758–763 (2002).
- Eder, W., Ludwig, W. & Huber, R. Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of Kebrit Deep, Red Sea. *Arch. Microbiol.* **172**, 213–218 (1999).
- Mouné, S., Caumette, P., Matheron, R. & Willison, J. C. Molecular sequence analysis of prokaryotic diversity in the anoxic sediments underlying cyanobacterial mats of two hypersaline ponds in Mediterranean salterns. *FEMS Microbiol. Ecol.* **44**, 117–130 (2003).
- Hinrichs, K.-U., Hayes, J. M., Sylva, S. P., Brewer, P. G. & DeLong, E. F. Methane-consuming archaeobacteria in marine sediments. *Nature* **398**, 802–805 (1999).
- Canfield, D. E. & Des Marais, D. J. Aerobic sulfate reduction in microbial mats. *Science* **251**, 1471–1473 (1991).
- Minz, D. *et al.* Unexpected population distribution in a microbial mat community: sulfate-reducing bacteria localized to the highly oxic chemocline in contrast to a eukaryotic preference for anoxia. *Appl. Environ. Microbiol.* **65**, 4659–4665 (1999).
- Eder, W., Jahnke, L. L., Schmidt, M. & Huber, R. Microbial diversity of the brine–seawater interface of the Kebrit Deep, Red Sea, studied via 16S rRNA

- gene sequences and cultivation methods. *Appl. Environ. Microbiol.* **67**, 3077–3085 (2001).
21. Ferrer, M. *et al.* Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem. Biol.* **12**, 895–904 (2005).
 22. Coolen, M. J. L. & Overmann, J. Functional exoenzymes as indicator of metabolically active bacteria in 124,000-year-old sapropel layers of the eastern Mediterranean Sea. *Appl. Environ. Microbiol.* **66**, 2589–2598 (2000).
 23. Coolen, M. J. L., Cypionka, H., Sass, A. M., Sass, H. & Overmann, J. Ongoing modifications of Mediterranean Pleistocene sapropels mediated by prokaryotes. *Science* **296**, 2407–2410 (2002).
 24. Simon, M. & Azam, F. Protein content and protein synthesis rates of planktonic marine bacteria. *Mar. Ecol. Prog. Ser.* **51**, 201–213 (1989).
 25. Takai, K. & Horikoshi, K. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl. Environ. Microbiol.* **66**, 5066–5072 (2000).
 26. Cardinale, M. *et al.* Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl. Environ. Microbiol.* **70**, 6147–6156 (2004).
 27. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, 1963).
 28. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
 29. Pearson, T. H. & Stanley, S. O. Comparative measurements of the redox potential of marine sediments as a rapid means of assessing the effect of organic pollution. *Mar. Biol.* **53**, 371–379 (1979).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the master and crew of the research vessel *Urania* for running the cruises; R. Knoop, S. Gusic, K. Bakker and P. Krystek for the geochemical analyses; and M. Cardinale and D. B. Nedwell for discussion. K.N.T. was supported by the Fonds der Chemischen Industrie. This study was supported by the European Commission's Sustainable Marine Ecosystem program, under the BioDeep project.

Author Information The newly determined 16S rRNA sequences have been submitted to the DDBJ/NCBI/GenBank database under accession numbers AM157647–AM157656, AY547745–AY547866 and DQ289238–DQ289401. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.D. (daniele.daffonchio@unimi.it).

Biodeep Scientific Party Jean Garcin¹, Boyd McKew², Peter N. Golyshin³, Nikolaos Lampadariou⁴, Paraskevi Polymenakou⁴, Daniele Calore⁵, Stefano Cenedese⁵, Fabio Zanon⁵ & Sven Hoog⁶

Affiliations for participants: ¹LMGEM, UMR 6117 CNRS Université de la Méditerranée, 13288 Marseille, France. ²Department of Biological Sciences, University of Essex, Colchester, Essex CO4 3SQ, UK. ³Division of Microbiology, GBF, 38122 Braunschweig, Germany. ⁴Institute of Marine Biology of Crete, 71003 Heraklion, Greece. ⁵Tecnomare SpA, ENI group, 30124 Venice, Italy. ⁶Technische Fachhochschule Berlin, University of Applied Science, 13353 Berlin, Germany.

LETTERS

Predator learning favours mimicry of a less-toxic model in poison frogs

Catherine R. Darst¹ & Molly E. Cummings¹

Batesian mimicry—resemblance of a toxic model by an edible mimic—depends on deceiving predators¹. Mimetic advantage is considered to be dependent on frequency because an increase in mimic abundance leads to breakdown of the warning signal^{2,3}. Where multiple toxic species are available, batesian polymorphism⁴ is predicted—that is, mimics diversify to match sympatric models. Despite the prevalence of batesian mimicry in nature⁵, batesian polymorphism is relatively rare⁶. Here we explore a poison-frog mimicry complex comprising two parapatric models and a geographically dimorphic mimic that shows monomorphism where models co-occur. Contrary to classical predictions, our toxicity assays, field observations and spectral reflectances show that mimics resemble the less-toxic and less-abundant model. We examine “stimulus generalization”⁷ as a mechanism for this non-intuitive result with learning experiments using naive avian predators and live poison frogs. We find that predators differ in avoidance generalization depending on toxicity of the model, conferring greater protection to mimics resembling the less-toxic model owing to overlap of generalized avoidance curves. Our work supports a mechanism of toxicity-dependent stimulus generalization⁸, revealing an additional solution for batesian mimicry where multiple models coexist.

In batesian mimicry, an edible species co-opts a warning signal from an unpalatable species to gain advantage through predator deception¹. If batesian mimics are too common, however, this advantage breaks down as predators learn to ignore the warning signal. Where more than one model species is available⁴, diversifying frequency-dependent selection predicts the evolution of

polymorphism in which mimics diverge in appearance to resemble sympatric models^{6,9,10}. Batesian polymorphism is suggested to distribute warning signal degradation over several defended model species, enabling the mimic to increase in abundance. Reported accounts of such mimetic polymorphism, however, are relatively rare⁶ and unknown in vertebrate mimicry systems^{11,12}.

Here we investigate a mimicry system that is inconsistent with the predictions of frequency dependence. We examine a poison-frog mimicry complex composed of two parapatric models and a geographically varying mimic (Fig. 1). The model Ecuadorian poison frogs *Epipedobates bilineatus* and *Epipedobates parvulus* share a similar warning signal of a bright red-spotted dorsum but differ in axilla and groin colouration (Fig. 1b). Their phylogenetically distant relative¹³, *Allobates zaparo*, is geographically dimorphic, matching each warning signal where models are parapatric (Fig. 1b). Where the two models co-occur, however, the mimic resembles only a single model (*E. bilineatus*; Figs 1 and 2). Here we use spectral reflectances, toxicity assays, field abundance measurements and predator learning experiments to investigate mechanisms that may be contributing to this pattern in nature.

Theoretical and empirical studies predict that coexistence of aposematic models may lead to (1) batesian polymorphism^{6,9,10}, (2) evolution of a mimic phenotype intermediate between model species^{14,15}, or (3) mimetic resemblance to the most highly abundant and/or noxious model^{16–19}. To test batesian mimicry predictions, we quantified patterns of mimicry, abundance, and toxicity of models and mimic in the zone of overlap. We assessed mimicry by degree of overlap between model and mimic using 95% confidence ellipses

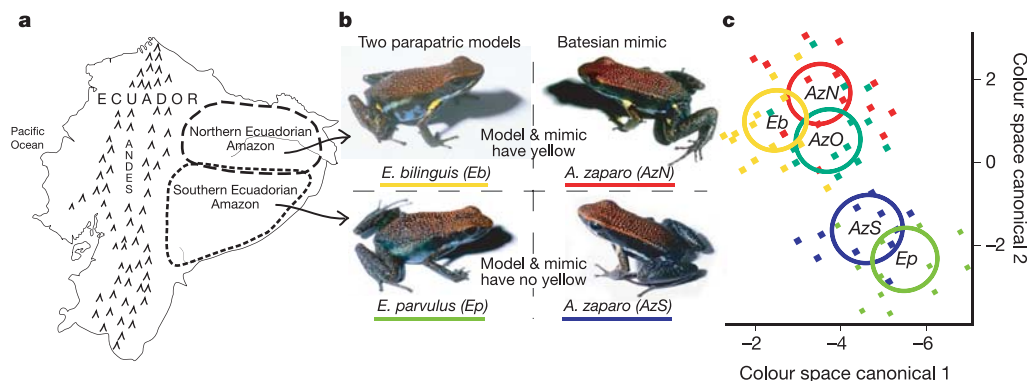


Figure 1 | Poison-frog mimicry complex and colour analyses. **a**, Geographic distribution of model and mimic species. **b**, Model and mimic warning signals. **c**, Discriminate functions plot with colour segments (radiance LS and MUV)²⁰ of ‘aposematic’ frog colours (red, yellow and black) from individuals’ head, dorsum, left and right axilla as covariates; and species and locality as categories. Mimicry was determined by overlap of model and

mimic 95% confidence ellipses around the multivariate centroid (Eb: *E. bilineatus*, $n = 25$; AzN: *A. zaparo* sympatric with *E. bilineatus* in north, $n = 15$; AzO: *A. zaparo* from model species’ zone of overlap, $n = 13$; AzS: *A. zaparo* sympatric with *E. parvulus* in south, $n = 13$; Ep: *E. parvulus*, $n = 28$).

¹Section of Integrative Biology, University of Texas, 1 University Station, Austin, Texas 78712, USA.

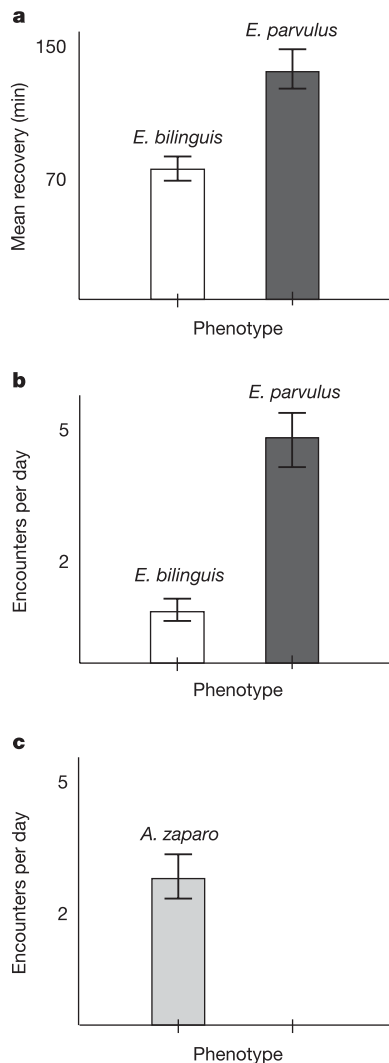


Figure 2 | Measured features of the poison-frog model-mimic system.

a, Relative toxicity of models. Shown is the mean recovery time (min) of mice after injection with different model skin extracts. **b**, Relative abundance of models (**b**) and mimic phenotype (**c**) where both models co-occur. Shown is the mean encounter rate per day. Data are mean $\pm$ s.e.m. The mimic in **c** assumed the *E. bilineus* phenotype.

computed from spectral reflectances²⁰ (Fig. 1c). The mimic, *A. zaparo*, shows significant divergence in colour pattern across its geographic range predicted by colour differences between model species (Fig. 1c). Where the two model species co-occur, however, the mimic's warning signal shows significant overlap with only *E. bilineus* (Fig. 1c). Thus, in contrast to predictions 1 and 2 for batesian mimics sympatric with multiple models, *A. zaparo* is neither polymorphic nor intermediate.

Applying prediction 3 to this poison-frog mimicry complex predicts that *A. zaparo* should mimic the more-toxic and/or abundant model where *E. parvulus* and *E. bilineus* co-occur. To test this prediction, we measured relative abundance as encounter rate across an 8-km transect on 10 consecutive days near the Río Arajuno, Napo Province, Ecuador. We found *E. parvulus* to be more abundant ($n = 43$ in total; mean $\pm$ s.e.m. frogs per day = 4.3 ± 0.62) and *E. bilineus* to be less abundant ($n = 10$ in total; 1.0 ± 0.26 frogs per day; Wilcoxon matched-pairs test, $Z = 2.716$; two-tail $P = 0.007$; Fig. 2b). We assessed the relative toxicity of the models and mimic using a standard protocol of frog skin extract subcutaneous injection into laboratory mice²¹. The time to recovery from injection of *E. parvulus* skin extract was significantly greater than the time to

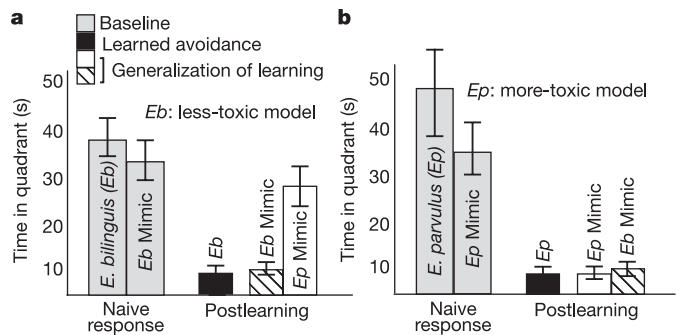


Figure 3 | Predator avoidance learning. Comparison of the chicks' baseline response time with postlearning time (mean $\pm$ s.e.m.) in the frog's test quadrant. **a**, Less-toxic model, *E. bilineus* (Eb), as learning stimulus. Chicks learned to avoid *E. bilineus* (baseline versus postlearning time: $Z = -2.207$, two-tail $P = 0.027$). Learned avoidance generalized to the *E. bilineus* mimic *A. zaparo* north ($Z = -2.201$, $P = 0.028$), but not to the *E. parvulus* (Ep) mimic *A. zaparo* south ($Z = -0.318$, $P = 0.75$). **b**, More-toxic model, *E. parvulus*, as learning stimulus. Chicks learned to avoid *E. parvulus* ($Z = -2.201$, $P = 0.028$). Learned avoidance generalized to the *E. parvulus* mimic *A. zaparo* south ($Z = -2.201$, $P = 0.028$), and also to the *E. bilineus* mimic *A. zaparo* north ($Z = -2.207$, $P = 0.027$). Data are mean $\pm$ s.e.m.

recovery from injection of *E. bilineus* skin extract ($n = 5$ mice per treatment; mean $\pm$ s.e.m. recovery time: 135.4 ± 9.31 min for *E. parvulus*, 79.0 ± 3.19 min for *E. bilineus*; $Z = 2.023$, two-tail $P = 0.043$; Fig. 2a). Injection of *A. zaparo* skin extract caused no adverse reaction (no difference among reactions from *A. zaparo* skin extract injections and saline control injections: 5.2 ± 1.8 min for *A. zaparo*; 5.1 ± 1.3 min for saline control). Thus, in contrast to prediction 3, *A. zaparo* mimics the less-abundant and less-toxic model, *E. bilineus*.

Mimics not only resemble the less-toxic model species in the overlap zone, they also outnumber these models significantly (2.6 ± 0.50 per day for *A. zaparo*, 1.0 ± 0.26 per day for *E. bilineus*; $Z_{n=10} = 2.09$; two-tail $P = 0.036$; Fig. 2b, c). To investigate why mimicry of a less-toxic and less-abundant model might be favoured by selection, we conducted predator-learning experiments to explore the classical^{7,8} psychological phenomenon of "stimulus generalization". Naive chicken predators were exposed to one of the model species in a series of learning trials, and then generalization of learned avoidance was assessed by subsequently exposing the educated predator to the precise mimic phenotype (found in sympatry with the learning stimulus) and the imperfect mimic phenotype (found in sympatry with the other model species). As predicted by single-model studies^{8,17,18}, we found that predator learning proceeded at a faster rate with the more-toxic model, *E. parvulus* (mean learning slope: 40.33 ± 8.11 for *E. parvulus*, 18.04 ± 7.4 for *E. bilineus*; $Z_{n=6} = 1.992$, two-tail $P = 0.046$). We tested mimic effectiveness (ability to deceive trained predators) and found that predators educated with either model (*E. bilineus* or *E. parvulus*) generalize learned avoidance, on sight, to their respective mimic phenotype of *A. zaparo* (Fig. 3; mean $\pm$ s.e.m. prelearning and postlearning time in quadrant: 25.83 ± 4.73 s and 5.33 ± 1.05 s, respectively, for *E. bilineus* mimic; 25.83 ± 4.17 s and 4.17 ± 1.54 s, respectively, for *E. parvulus* mimic; $Z_{n=6} = 2.201$, two-tail $P = 0.028$), providing empirical evidence for batesian mimicry in dendrobatid frogs.

We further examined how broadly generalization of avoidance extends or how imperfect a mimic can be and still gain protection from predators educated with a specific model. Although precise mimics enjoyed equal protection regardless of the model species used for learning, imperfect mimics did not. Generalization of learned avoidance to the imperfect mimic differed depending on the toxicity of the model learning stimulus (Fig. 3; mean $\pm$ s.e.m. postlearning time with imperfect mimic: 6.67 ± 1.05 s for *E. parvulus* as learning

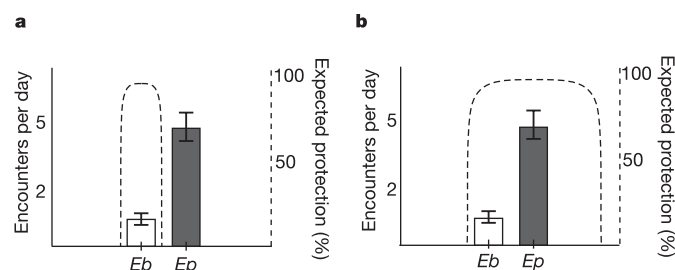


Figure 4 | Generalized avoidance curves. Broken lines represent expected protection for each phenotype estimated using the predator learning data in Fig. 3. Estimates of protection assume fully trained predators in the wild. **a**, Less-toxic model, *E. bilinguis*, as learning stimulus. Learned avoidance does not generalize beyond the warning signal with which predators were trained. **b**, More-toxic model, *E. parvulus*, as learning stimulus. Learned avoidance generalizes to the mimic of both *E. parvulus* and *E. bilinguis*. Thus, *A. zaparo* individuals resembling less-toxic *E. bilinguis* gain a selective advantage no matter which model is the avoidance learning stimulus. Data are mean $\pm$ s.e.m.

stimulus, 26.67 ± 4.41 s for *E. bilinguis* as learning stimulus; $Z_{n=6} = 2.207$; two-tail $P = 0.027$). Predators educated with the less-toxic model, *E. bilinguis*, did not generalize learned avoidance to the mimic of *E. parvulus* (Fig. 3a; baseline time in quadrant: 28.34 ± 4.41 s; postlearning time with imperfect mimic: 26.67 ± 4.41 s; $Z_{n=6} = 0.318$; two-tail $P = 0.75$). By contrast, predators educated with the more-toxic model, *E. parvulus*, did generalize learned avoidance to the imperfect mimic, the mimic of *E. bilinguis* (Fig. 3b; baseline time: 49.17 ± 9.17 s; postlearning time with imperfect mimic: 6.67 ± 1.05 s; $Z_{n=6} = 2.201$; two-tail $P = 0.028$). Thus, the stimulus generalization gradient is broader when avoidance is learned on the more-toxic model (avoidance generalizes to both mimic phenotypes) and, in contrast, the stimulus generalization gradient is more narrow when avoidance is learned on the less-toxic model (avoidance generalizes to only the precise mimic phenotype^{8,15}; Fig. 4).

The relative selective advantage gained by either mimic phenotype in the zone of model species overlap is dependent on the penalty to the predator from the particular model being mimicked. Learned avoidance from experience with the more-toxic model will generalize to either mimic phenotype; both mimic phenotypes receive protection if the predator has undergone avoidance learning with more-toxic *E. parvulus* (Fig. 4b). Learned avoidance from experience with less-toxic *E. bilinguis*, however, will generalize only to the precise mimic of *E. bilinguis* (Fig. 4a). In the zone of model species overlap, therefore, mimics of *E. parvulus* only receive protection generated by *E. parvulus*, whereas mimics of *E. bilinguis* receive benefits generated by both models.

An alternative explanation for the apparent mimicry mismatch, wherein the mimic resembles the less-toxic and less-abundant model in the overlap zone, may be recent model range expansion (*E. parvulus*) or contraction (*E. bilinguis*) in this region. If the range of *E. parvulus* recently expanded north, or if *E. bilinguis* populations recently shrank in the overlap zone, then we may be capturing this species complex in an evolutionary lag snapshot—in which the mimic (*A. zaparo*) has not had enough ‘time’ to show perfect mimicry to the more-abundant and more-toxic model. Although no range transformation data are available to test this possibility conclusively, it does not rule out the idea that toxicity-dependent generalized avoidance may maintain the current imbalance between mimic and model.

By mimicking the less-toxic model (rather than mimetic polymorphism, an intermediate mimic phenotype, or mimicking the most toxic and/or numerous model), the increased predation risk accrued by an increased abundance of batesian mimic individuals is

spread over both defended model species, enabling the mimic to increase in abundance. This non-intuitive result is driven by toxicity-dependent generalization of learned avoidance: predators that learn on the more-toxic model will generalize avoidance to the less-toxic model’s mimic, whereas predators that learn on the less-toxic model show no generalization beyond this precise warning signal^{8,15}. Thus, a mimic of the less-toxic model can enjoy near complete protection from educated predators regardless of which model was used for avoidance learning. We have presented strong evidence suggesting that the selective force influencing *A. zaparo*’s resemblance of the less-toxic and less-abundant model, *E. bilinguis*, is stimulus-controlled predator generalization of learned avoidance. Our work therefore provides an adaptive hypothesis based on the classical psychological phenomenon of stimulus generalization^{7,8}, which may help to explain the paucity of batesian polymorphism examples, and reveals a monomorphic evolutionary solution to the problem of batesian abundance.

METHODS

Collection and abundance estimates. Fieldwork was conducted in Amazonian lowland rainforest, between January and May in 2003–2005. In February 2004, we measured poison-frog encounter rates along a ~8-km transect trail for 10 consecutive days in the overlap zone, Río Arajuno (~3 km southwest of San Pedro), Napo, Ecuador, of the models. For reflectance measurements and predation experiments, we collected live frogs from Estación Biológica Jatun Sacha, Napo (*Allobates zaparo* and *Epipedobates bilinguis*); Río Arajuno, Napo (*A. zaparo*, *E. bilinguis*, and *E. parvulus*); and Santiago, Morona-Santiago (*A. zaparo* and *E. parvulus*). For predation experiments, we collected brown, nontoxic *Colostethus awa* from western Ecuadorian cloudforest at Union del Tuachi, Pichincha. Taxonomy was as described²².

Colour analyses. Ninety-four frogs were collected and transported to Museo de Zoología, Universidad Católica del Ecuador (Fig. 1c). Spectral reflectances were measured with an Ocean Optics PS2000 spectrometer, DT-1000 full-spectrum light source, Spectralon white standard and reflectance probe (R400-7) at a 2-mm distance from seven body regions: head, dorsum, axilla, groin, vocal sac, flanks and ventor (two measures per region). We collected leaf-litter background reflectances (Jatun Sacha, six; Río Arajuno, seven; Santiago, seven). Forty-five habitat spectral irradiance measurements were collected at 0900 on 9 d with the PS2000 and cosine collector. Frog and background radiance estimates were computed as the product of spectral reflectances and average habitat irradiance spectrum.

To compare radiance measurements, independent of the visual system, we used a segments classification method²⁰. Radiance spectra were divided into four bandwidths (ultraviolet, 300–399 nm; short, 400–499 nm; middle, 500–599 nm; long, 600–699 nm), normalized by total intensity, and evaluated in a two-dimensional space by orthogonal axes representing hypothetical opponency processes (LS, long–short; MUV, middle–ultraviolet). We computed composite euclidean distances²⁰, D_{comp} , representing distance in colour space between frog and leaf-litter background. Whole-body colouration measures were similar between model species (D_{comp} : 22.18 ± 3.88 for *E. parvulus*, 20.24 ± 5.30 for *E. bilinguis*; $t = 0.456$, two-tail $P = 0.664$). To evaluate mimicry, we used multivariate discriminate functions analyses of warning coloured segments in JMP^{23,24} (Fig. 1c).

Toxicity assays. Five frogs from each species were killed and skinned as described²⁵. Methanol extracts from individual frogs were evaporated and resuspended in sterile saline. Resultant single-skin extracts were subcutaneously injected in four treatments, each given to five mice²¹ ($n = 20$ mice, IACUC 03110501), as follows: *E. bilinguis*, *E. parvulus*, *A. zaparo* or saline control. Sleeping behaviour was the baseline for toxicity assays. Mice were awakened with the injection and the time to complete recovery (return to sleep) was recorded. Recovery time was used to estimate the degree of toxicity.

Predator learning experiments. Although few data exist, birds may be potential predators of poison frogs²⁶. Thus, in Ecuador we conducted a series of learning experiments using ~1-month-old domestic chickens (*Gallus gallus domesticus*) as naive, model predators²⁷ and wild-caught dendrobatid frogs (models, *E. bilinguis* and *E. parvulus*; mimic, *A. zaparo*). Birds were tested individually in a 1-m² dirt-floor test arena of four 50-cm² quadrants, outside under natural lighting conditions. Chickens were fed chicken mash and cracked corn twice daily and water *ad libitum*. We assessed mimic palatability by presenting six naive chickens with an *A. zaparo* (three northern and three southern *A. zaparo*). Naive chickens readily ate both *A. zaparo* and control frogs (*C. awa*).

We had two experimental groups (six chicks each), which differed in learning

stimulus (*E. parvulus* or *E. bilinguis*) in eight learning trials (IACUC 04071901). Learning trials consisted of presenting a chick with a learning stimulus under a glass dome for 1 min or until the chick pecked the dome. The dome was then removed and latency to approach the stimulus was recorded for up to 2 min or until a sampling event. In a typical sampling event, chicks grabbed the frog in their beaks and spat the frog out. Only one chick ingested a poison frog (*E. bilinguis*); it died 3 d later and its data were removed. All other chicks tasted and released the frog; most frogs survived the sampling event. We defined the learning rate as the slope (latency to peck divided by number of trials) until complete avoidance (no subsequent sampling in further trials). Control frogs were presented to chicks after trials 2 and 6 to assure that chicks were still motivated to eat frogs.

After training was complete, learning and learning generalization were assessed in choice experiments that paired the control frog with one of three brightly coloured dendrobatid frogs: toxic model learning stimulus (learned avoidance); precise mimic of learning stimulus (learning generalization); and imperfect mimic of learning stimulus (degree of generalization). Chicks were presented with both the brightly coloured frog and a control frog each under a glass dome for 2 min, and the time spent in the test arena quadrant of each dome was recorded. Frog placement in the test arena was randomized across trials. We assessed learned avoidance and generalization of learned avoidance by comparing prelearning (baseline) and postlearning time spent by chicks in the brightly coloured frog's test arena.

Received 31 August; accepted 4 October 2005.

- Bates, H. W. Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. *Trans. Linn. Soc. Lond.* **23**, 495–566 (1862).
- Fisher, R. A. *Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
- Brower, L. P. & Brower, J. V. Z. The relative abundance of model and mimic butterflies in natural populations of the *Battus philenor* mimicry complex. *Ecology* **43**, 154–158 (1962).
- Wallace, A. R. On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan region. *Trans. Linn. Soc. Lond.* **25**, 1–71 (1865).
- Edmunds, M. *Defence in Animals* (Longman, Harlow, 1974).
- Joron, M. & Mallet, J. L. B. Diversity in mimicry: paradox or paradigm? *Trends Ecol. Evol.* **13**, 461–466 (1998).
- Pavlov, I. P. *Conditioned Reflexes* (translated by Anrep, G. V.) (Univ. Press, London, 1927).
- Duncan, C. J. & Sheppard, P. M. Sensory discrimination and its role in the evolution of Batesian mimicry. *Behaviour* **24**, 269–282 (1965).
- Ford, E. B. *Ecological Genetics* 3rd edn (Chapman and Hall, London, 1971).
- Turner, J. R. G. The evolutionary dynamics of Batesian and Müllerian mimicry: similarities and differences. *Ecol. Entomol.* **12**, 81–95 (1987).
- Brodie, E. D. Jr & Brodie, E. D. III Differential avoidance of mimetic salamanders by free-ranging birds. *Science* **208**, 181–182 (1980).
- Greene, H. W. & McDiarmid, R. W. Coral snake mimicry: does it occur? *Science* **213**, 1207–1212 (1981).
- Santos, J. C., Coloma, L. A. & Cannatella, D. C. Multiple, recurring origins of aposematism and diet specialization in poison frogs. *Proc. Natl Acad. Sci. USA* **100**, 12792–12797 (2003).
- Edmunds, M. Why are there good mimics and poor mimics? *Biol. J. Linn. Soc.* **70**, 459–466 (2000).
- Sherratt, T. N. The evolution of imperfect mimicry. *Behav. Ecol.* **13**, 821–826 (2002).
- Brower, J. V. Z. Experimental studies of mimicry. IV. The reactions of starlings to the different proportions of models and mimics. *Am. Nat.* **94**, 271–282 (1960).
- Goodale, M. A. & Sneddon, I. The effect of distastefulness of the model on the predation of artificial Batesian mimics. *Anim. Behav.* **25**, 660–665 (1977).
- Lindström, L., Alatalo, R. V. & Mappes, J. Imperfect Batesian mimicry—the effects of the frequency and the distastefulness of the model. *Proc. R. Soc. Lond. B* **264**, 149–153 (1997).
- Johnstone, R. A. The evolution of inaccurate mimics. *Nature* **418**, 524–526 (2002).
- Endler, J. A. On the measurement and classification of colour in studies of animal colour patterns. *Biol. J. Linn. Soc.* **41**, 315–352 (1990).
- Daly, J. W., Garraffo, H. M. & Spande, T. F. in *Alkaloids: Chemical and Biological Perspectives* vol. 13 (ed. Pelletier, S. W.) 1–161 (Pergamon, New York, 1999).
- Frost, D. R. *Amphibian Species of the World 3.0: an Online Reference* (<http://research.amnh.org/herpetology/amphibia/index.html>) (American Museum of Natural History, New York, 2004).
- SAS Institute. *JMP Statistics and Graphics Guide* version 4 (SAS Institute, Cary, NC, 2000).
- Sokal, R. R. & Rohlf, F. J. *Biometry: the Principles and Practice of Statistics in Biological Research* 2nd edn (W. H. Freeman and Co., San Francisco, 1981).
- Darst, C. R., Menéndez-Guerrero, P. A., Coloma, L. A. & Cannatella, D. C. Evolution of dietary specialization and chemical defense in poison frogs (Dendrobatidae): a comparative analysis. *Am. Nat.* **165**, 56–69 (2005).
- Master, T. L. *Dendrobates auratus* (black and green poison dart frog): predation. *Herp. Rev.* **29**, 164–165 (1998).
- Osorio, D., Vorobyev, M. & Jones, C. D. Colour vision of domestic chicks. *J. Exp. Biol.* **202**, 2951–2959 (1999).

Acknowledgements We thank L. Coloma, J. C. Santos, and S. Ron for discussions on poison frogs; D. Cannatella for frog photos in Fig. 1, assistance with assays and discussion; M. Domjan for advice on predator learning experiments; E. Tapia, S. Padilla, M. Bustamante, P. Menéndez-Guerrero and D. Paucar for assistance in the field; J. Cassaday for assistance with toxicity assays; and M. Ryan for comments on the manuscript. This work was supported by University of Texas at Austin EEB graduate program fellowships, a UT Continuing Fellowship, and the Explorer's Club Exploration Fund.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.R.D. (catdarst@mail.utexas.edu).

LETTERS

Global tests of biodiversity concordance and the importance of endemism

John F. Lamoreux¹, John C. Morrison², Taylor H. Ricketts², David M. Olson³, Eric Dinerstein², Meghan W. McKnight⁴ & Herman H. Shugart¹

Understanding patterns of biodiversity distribution is essential to conservation strategies¹, but severe data constraints make surrogate measures necessary^{2–4}. For this reason, many studies have tested the performance of terrestrial vertebrates as surrogates for overall species diversity, but these tests have typically been limited to a single taxon or region^{3–10}. Here we show that global patterns of richness are highly correlated among amphibians, reptiles, birds and mammals, as are endemism patterns. Furthermore, we demonstrate that although the correlation between global richness and endemism is low, aggregate regions selected for high levels of endemism capture significantly more species than expected by chance. Although areas high in endemism have long been targeted for the protection of narrow-ranging species^{11,12}, our findings provide evidence that endemism is also a useful surrogate for the conservation of all terrestrial vertebrates.

One of the challenges to preserving the Earth's biota is that species are unevenly distributed¹. Confronted with the continuing extinction crisis, conservation strategies often focus either on areas with high species richness to maximize the number of species covered, or on areas that contain large numbers of endemic species (species found nowhere else)^{11–14} (Fig. 1a, b). The vast majority of species, however, have yet to be named, and information regarding their ranges (as well as the geographic ranges of many described species) is lacking^{15,16}. Vertebrates, being relatively well-known, are frequently used to represent all biodiversity, but analyses of cross-taxa congruence often show little overlap^{5–7} and have thus lowered confidence in the use of surrogates. This lack of concordance could be a consequence of limitations in the taxonomic breadth or geographic extent of previous studies^{3–10}.

Here we use a uniquely comprehensive data set of terrestrial vertebrate distributions to evaluate global concordance in diversity patterns among four classes: amphibians, reptiles, birds and mammals. We recorded the presence or absence of 26,452 species according to 799 terrestrial ecoregions of the world¹⁷ (data available at <http://www.worldwildlife.org/wildfinder>). These ecoregions serve as the basis of World Wildlife Fund's conservation planning¹⁴, The Nature Conservancy's international efforts¹⁸, and the delineation of Conservation International's hotspots¹⁹ and high biodiversity wilderness areas²⁰. Because ecoregions are created with species assemblages in mind, they are more useful units for comparing species data than are national lists²¹.

Richness is perhaps the most common measure of species diversity because it is relatively easy to compile and it implies that large numbers of species can be conserved in small areas^{1,9,13}. We calculated the correlation between the proportional richness corrected for ecoregion area (hereafter termed 'richness'; see Methods) of each class and the other vertebrates (an index combining the three

remaining classes). Correlations between richness within a class and the richness of the other classes were strong, positive and significant for amphibians, birds and mammals (Pearson correlation coefficients, which are used throughout, are 0.591, 0.715 and 0.668, respectively; $P < 0.01$) (Table 1). A positive and significant correlation between reptiles and other vertebrates also existed, but was more moderate (0.380, $P < 0.01$) (Table 1). Ongoing debates over the causes of coincident patterns of global biodiversity include hypotheses regarding climate, geologic history, and a statistical consequence of differences in range size among species. All have explanatory merit, and a combination of factors seems likely²².

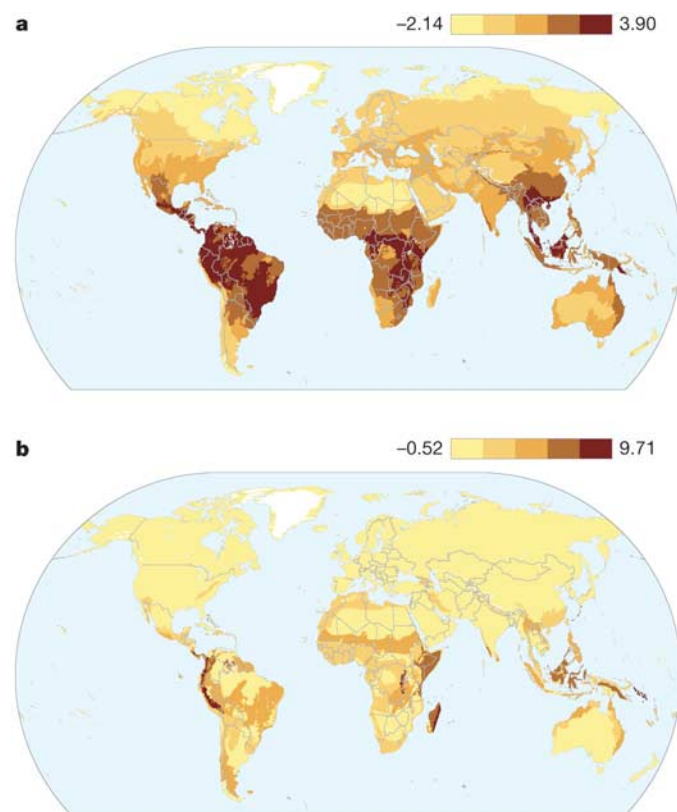


Figure 1 | Terrestrial vertebrate diversity by ecoregion. These proportional indices combine the four terrestrial vertebrate classes and adjust for ecoregion area. Each scale bar of five colours represents relative levels of diversity from low (light) to high (dark). **a**, Species richness. **b**, Species endemism.

¹Department of Environmental Sciences, University of Virginia, 291 McCormick Road, Charlottesville, Virginia 22904, USA. ²World Wildlife Fund-US, 1250 24th Street NW, Washington DC 20037, USA. ³Wildlife Conservation Society-South Pacific, 11 Ma'afu Street, Suva, Fiji. ⁴Curriculum in Ecology, Miller Hall, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

Regardless of the underlying mechanisms, our results confirm that patterns of species richness among terrestrial vertebrates are broadly concordant.

Endemic species are another important target of global conservation efforts^{11,12}. These species, often having small populations and few sites for conservation intervention, are inherently vulnerable to extinction²³. As for species richness, a smaller amount of total area will need to be conserved if endemism patterns among taxa are correlated. We found that proportional endemism (hereafter termed 'endemism'; see Methods) within each vertebrate class showed significantly positive correlations with endemism of the remaining three classes (Table 1). Endemism of reptiles (0.587, $P < 0.01$) and birds (0.612, $P < 0.01$) were more strongly related to their respective three-class index than endemism of mammals (0.490, $P < 0.01$) and amphibians (0.503, $P < 0.01$). These findings indicate that endemism within a well-documented group may be useful for guiding conservation decisions regarding overall endemism. Of course, correlations among classes only explain a portion of the observed variance in endemism among ecoregions. Therefore, conservation strategies based on one taxon will have to be supplemented with specific information for other groups in order to capture all endemic species.

Studies of congruence between richness and endemism have been inconclusive at regional scales^{8,24}, and the one previous global test showed low congruence within birds⁸. Our analysis found no meaningful correlation between richness and endemism within any of the four terrestrial vertebrate classes (correlation coefficients -0.099 to 0.096) or for vertebrates overall (-0.025) (Table 1), confirming that global patterns of these diversity measures are not spatially concordant. Thus, global conservation priorities based on richness alone will overlook many endemic species.

Although global correlations are suggestive of concordant diversity patterns, the question most relevant to conservation decisions is whether a specific set of ecoregions selected for one measure will represent non-target species^{3,4,25}. Because endemic species are particularly important, we test how efficiently ecoregions chosen based on the level of endemism capture all species (Fig. 2a). After only 10% of the terrestrial land area is selected (an arbitrary cutoff⁴), the aggregated set of ecoregions contain 56.5% of the world's terrestrial vertebrate endemics. Of more interest here, these ecoregions capture 61.6% of all vertebrates, significantly higher than expected through random ecoregion selection (1,000 randomizations, $47.5 \pm 2.2\%$, $P < 0.01$) (Supplementary Table 1 and Supplementary Fig. 1). The numbers of both endemics and all species captured continue to rise more steeply than expected, reaching an asymptote at roughly 50% of the Earth's surface (Fig. 2a). However, the actual area is much smaller than indicated by the analysis, simply because no ecoregion is pristine and most have large areas that are now of limited use for biodiversity. Selecting ecoregions on the basis of the endemism of a single class sometimes performs even better than using all four. For example, 10% of the world's land area chosen on the basis of bird endemism alone captures 59.6% of all vertebrate endemics; and on the basis of amphibian and reptile endemism, 71.0% and 72.0%,

respectively, of total vertebrate species are captured (Supplementary Table 2). Our findings demonstrate that, although broad correlations between richness and endemism are weak, priority sets based on endemic species contain large numbers of total species. This result is probably due to turnover in species composition among areas of high endemism (that is, these areas are highly complementary in terms of non-endemic species²⁶) and suggests that endemism is particularly useful for conservation prioritization.

Although our work furthers the understanding of how species diversity patterns can inform conservation priorities at a global scale, we must point to several important caveats. The distribution patterns we report only apply to vertebrates and might not hold for more species-rich taxa such as plants, invertebrates and fungi. Biodiversity of vertebrates also encompasses aspects of population and genetic diversity that are missed when using species as the sole unit of measurement. Furthermore, methods for setting conservation priorities are complex and should consider not just the number of endemics or total species present, but also degree of threat^{7,11}, population viability¹⁸, ecological and evolutionary processes^{14,18}, and economic costs and benefits of conservation²⁷. Comprehensive conservation strategies will require efforts at multiple scales to ensure the long-term survival of biodiversity in a region^{2,18}. Using endemism along with other factors to identify global priorities helps to focus these conservation efforts on critical regions^{11,12,14}, where on-the-ground efforts will yield the greatest payoffs for biodiversity.

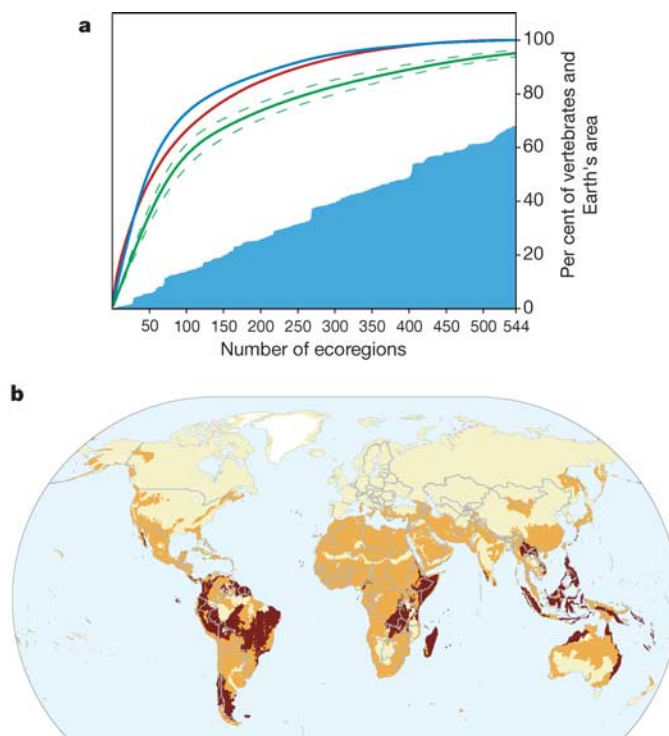


Figure 2 | Accumulation of species captured when selecting ecoregions on the basis of endemism. **a**, The percentage of endemic species (red line) and all species (dark blue line) rise rapidly as the per cent of ecoregion area selected (blue shading) increases. The green line represents the total per cent of species captured when ecoregions are selected at random (mean of 1,000 random sets). Species accumulate significantly faster than expected by chance (lying above a 95% confidence interval, dashed line), and nearly all terrestrial vertebrates are found within ecoregions that represent 50% of the Earth's surface. **b**, Ecoregions representing 10% of the Earth's terrestrial surface (dark brown) that capture 56.5% of vertebrate endemics and 61.6% of all species. Also depicted are the additional ecoregions required to represent 50% of the Earth (orange); together, the two sets of ecoregions contain nearly all species of terrestrial vertebrates.

Table 1 | Pearson correlation coefficients of terrestrial vertebrate diversity measures

	Amphibians	Reptiles	Birds	Mammals	Four Classes
Richness†	0.591**	0.380**	0.715**	0.668**	
Endemism‡	0.503**	0.587**	0.612**	0.490**	
Richness × Endemism§	0.096**	0.085**	-0.068	-0.099	-0.025

* $P < 0.05$; ** $P < 0.01$.

†Correlation between class richness and a richness index of the three remaining classes.

‡Correlation between class endemism and an endemism index of the three remaining classes.

§Correlation between richness and endemism within each class, and of the four classes combined.

METHODS

Database. The database used for these analyses contains presence/absence data for the world's terrestrial amphibians ($n = 4,797$), reptiles ($n = 7,483$), birds ($n = 9,470$) and mammals ($n = 4,702$) by ecoregion. These data are based on the natural, historic ranges of extant species, such that species introduced, present as human commensals, vagrants or passage migrants were not recorded.

The terrestrial ecoregions are those delineated in ref. 17. Mangrove ecoregions were excluded from analysis because the land area of these regions is invariably small; hence species lists compiled from overlaying range maps inflates species totals. A further nine ecoregions were excluded from the analysis because of lack of data, but we are confident that these omissions do not affect overall distribution patterns (land area of the world's ecoregions is 134,735,751 km², land area of the nine excluded ecoregions is 236,100 km² or 0.175%). Because the large interior portions of Greenland and Antarctica contain no terrestrial vertebrates and are not mapped as ecoregions¹⁷, they too were excluded from the analysis as well as from calculations of Earth's land area.

Analyses. Class richness was tallied for each ecoregion and divided by the total number of species in the database for that class¹⁰ (or the total number within biomes when the data were regressed separately; see below). This proportional species richness allowed calculations to be comparable between taxa without a single species group overwhelming the others:

$$\text{Index}_{(e)} = \frac{\sum_{i=1}^n G_{i(e)}}{\sum_{i=1}^n G_{i(t)}}$$

where e is the richness index for each ecoregion, n is the number of species groups used in the index, $G_{i(e)}$ is the number of species in group i per ecoregion, and $G_{i(t)}$ is the total number of species of group i . Endemism indices were calculated in the same manner. An adjusted richness index was used for comparisons between overall richness and endemism, whereby the richness totals included only non-endemic species so that endemics were not part of both comparative sets²⁴.

We regressed each index against area (both variables were log₁₀-transformed) in order to reduce the influence of ecoregion size on the indices. Because the relationship between species richness and area varied by biome, we performed regressions for each biome separately, and used the combined residuals of each index for subsequent analyses. There was, however, no significant relationship between the endemism indices and area, and we therefore used the original indices for endemism. There was also no effect of latitude after we accounted for area.

Pearson correlations were used for all analyses. We tested for statistical significance using a geographically restricted randomization approach with 10,000 iterations. Randomizations were constrained by both biome and biogeographic realm to reduce Type I errors owing to spatial autocorrelation²⁸. This method is more stringent than unrestricted randomizations or randomizations restricted to either biomes or realms²⁸. We used the following standards to interpret correlation coefficients: a large correlation coefficient was approximately 0.50 and higher, moderate correlations were around 0.30, and small correlations were about 0.10 (refs 29, 30).

To determine whether the percentage of species captured when selecting ecoregions for endemism was greater than by chance alone, we selected 1,000 random sets of ecoregions for every 10% interval in area, up to 67.9% (the percentage at which all species were captured). Each iteration selected the same number of ecoregions as in the set chosen for endemism at the corresponding interval (see Supplementary Information). Note that because ecoregions vary in size, the percentages of Earth's land surface are slightly less than the 10% breaks. For example, ecoregions selected for the most endemics within 10% of the global land area actually make up 9.29% (always favouring the more conservative number).

Received 27 June; accepted 29 September 2005.

Published online 28 December 2005.

- Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
- Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405**, 243–253 (2000).
- Howard, P. C. *et al.* Complementarity and the use of indicator groups for reserve selection in Uganda. *Nature* **394**, 472–475 (1998).
- Moore, J. L. *et al.* Performance of sub-Saharan vertebrates as indicator groups for identifying priority areas for conservation. *Conserv. Biol.* **17**, 207–218 (2003).
- Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C. & Gibbons, D. W. Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337 (1993).

- van Jaarsveld, A. S. *et al.* Biodiversity assessment and conservation strategies. *Science* **279**, 2106–2108 (1998).
- Dobson, A. P., Rodriguez, J. P., Roberts, W. M. & Wilcove, D. S. Geographic distribution of endangered species in the United States. *Science* **275**, 550–553 (1997).
- Orme, C. D. L. *et al.* Global hotspots of species richness are not congruent with endemism or threat. *Nature* **436**, 1016–1019 (2005).
- Brooks, T. *et al.* Toward a blueprint for conservation in Africa. *Bioscience* **51**, 613–624 (2001).
- Ricketts, T. H., Dinerstein, E., Olson, D. M. & Loucks, C. Who's where in North America? *Bioscience* **49**, 369–381 (1999).
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- Stattersfield, A. J., Crosby, M. J., Long, A. J. & Wege, D. C. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation* (BirdLife International, Cambridge, 1998).
- Mittermeier, R. A., Robles Gil, P. & Mittermeier, C. G. *Megadiversity: Earth's Biologically Wealthiest Nations* (CEMEX, Mexico City, 1997).
- Olson, D. M. & Dinerstein, E. The global 200: A representation approach to conserving the Earth's most biologically valuable ecoregions. *Conserv. Biol.* **12**, 502–515 (1998).
- Raven, P. H. & Wilson, E. O. A fifty-year plan for biodiversity surveys. *Science* **258**, 1099–1100 (1992).
- Millennium Ecosystem Assessment. *Ecosystems and Human Well-being: Biodiversity Synthesis* (World Resources Institute, Washington DC, 2005).
- Olson, D. M. *et al.* Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* **51**, 933–938 (2001).
- Groves, C. R. *Drafting a Conservation Blueprint: A Practitioner's Guide to Planning for Biodiversity* (Island Press, Washington DC, 2003).
- Mittermeier, R. A. *et al.* Hotspots Revisited (CEMEX, Mexico City, 2004).
- Mittermeier, R. A. *et al.* Wilderness and biodiversity conservation. *Proc. Natl Acad. Sci. USA* **100**, 10309–10313 (2003).
- Pimm, S. L. Seeing both the woods and the trees. *Nature* **402**, 853–854 (1999).
- Pimm, S. L. & Brown, J. H. Domains of diversity. *Science* **304**, 831–833 (2004).
- Gaston, K. J. Rarity as double jeopardy. *Nature* **394**, 229–230 (1998).
- Ricketts, T. H. Aligning conservation goals: Are patterns of species richness and endemism concordant at regional scales? *Anim. Biodivers. Conserv.* **24**, 91–99 (2001).
- Balmford, A. On hotspots and the use of indicators for reserve selection. *Trends Ecol. Evol.* **13**, 409 (1998).
- Williams, P. H. in *Conservation in a Changing World* (eds Mace, G. M., Balmford, A. & Ginsberg, J. R.) 211–249 (Cambridge Univ. Press, Cambridge, 1998).
- Balmford, A. *et al.* Economic reasons for conserving wild nature. *Science* **297**, 950–953 (2002).
- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* 2nd edn (Chapman and Hall, London, 1997).
- Aron, A. & Aron, E. N. *Statistics for Psychology* 3rd edn (Prentice Hall, Upper Saddle River, New Jersey, 2003).
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Erlbaum, Hillsdale, New Jersey, 1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by World Wildlife Fund's Living Planet Campaign with support from J. Leape. We thank P. Ehrlich, K. Gaston and H. Possingham for comments; and the following for expertise, unpublished data and technical assistance: A. Allison, T. Allnutt, S. Anderson, L. Aquino, P. Beresford, G. Brennan, D. Broadley, R. Brown, C. Burdette, N. Burgess, L. Burrus, J. Campbell, F. Castro, L. Cayet, E. Colijn, L. Coloma, R. Crombie, I. Das, I. de La Riva, A. Diesmos, W. Duellman, K. Ernst, M. Evans, L. Fitzgerald, J. Fjelds , J. Formas, D. Frost, S. Galster, A. Gascoigne, T. Grant, M. Hadjichristoforou, J. D'Amico Hales, L. Hansen, R. Heyer, R. Hoffmann, M. Hoogmoed, R. Ibanez, J. Icochea, D. Iskandar, N. Jorge da Silva, G. K hler, C. Kurtonur, E. La Marca, J. Langone, E. Lavilla, F. Lobo, S. L tters, C. Loucks, S. Madge, O. Marques, M. Martins, J. McCranie, R. McDiarmid, R. Montero, J. Ortiz, S. Palminteri, J. Patterson, J. Pilgrim, G. Pinto, J. Porter, G. Powell, C. Rahbek, S. Reichle, E. Richard, G. Rodda, J. Savage, E. Schaub, J. Schipper, N. Scott, G. Scrocchi, W. Sechrest, H. Strand, S. Stuart, M. Taye, R. Tizard, D. Turney, D. Urban, P. Uetz, E. Underwood, P. van Dijk, O. Villela, J. Weiss, W. Wetters, L. Wilson, Y. Xie, A. Yanosky, E. Zhao and G. Zug. We also thank the Species Survival Commission of IUCN, The Nature Conservancy and the Zoological Museum of the University of Copenhagen for data contributions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.F.L. (lamoreux@virginia.edu).

V1 spinal neurons regulate the speed of vertebrate locomotor outputs

Simon Gosgnach^{1*}, Guillermo M. Lanuza^{1*}, Simon J. B. Butt^{3†}, Harald Saueressig¹, Ying Zhang¹, Tomoko Velasquez¹, Dieter Riethmacher⁴, Edward M. Callaway², Ole Kiehn³ & Martyn Goulding¹

The neuronal networks that generate vertebrate movements such as walking and swimming are embedded in the spinal cord^{1–3}. These networks, which are referred to as central pattern generators (CPGs), are ideal systems for determining how ensembles of neurons generate simple behavioural outputs. In spite of efforts to address the organization of the locomotor CPG in walking animals^{2,4–6}, little is known about the identity and function of the spinal interneuron cell types that contribute to these locomotor networks. Here we use four complementary genetic approaches to directly address the function of mouse V1 neurons, a class of local circuit inhibitory interneurons that selectively express the transcription factor Engrailed1. Our results show that V1 neurons shape motor outputs during locomotion and are required for generating ‘fast’ motor bursting. These findings outline an important role for inhibition in regulating the frequency of the locomotor CPG rhythm, and also suggest that V1 neurons may have an evolutionarily conserved role in controlling the speed of vertebrate locomotor movements.

Several genetically defined classes of neurons have been identified in the developing spinal cord⁷, including a class of ipsilaterally

projecting inhibitory neurons that innervate motor neurons, the Engrailed1 (En1)-expressing V1 neurons^{8–11}. To assess the function of V1 neurons in the locomotor CPG, we used two mouse models that have selective loss of the En1-expressing V1 neuronal population (Fig. 1). This was achieved through the altered specification of V1 neurons in *Pax6*-knockout (*Pax6*^{−/−}) mice and by the selective ablation of these neurons in *En1*^{Cre}; *R26-lacZ*^{fllox}/*DTA* (*En1-DTA*; ref. 12) mice. At embryonic day (E)12.5, a marked reduction in En1-positive V1 neuron cell numbers was apparent in spinal cords isolated from *Pax6*^{−/−} and *En1-DTA* embryos (Fig. 1a, e). This was confirmed using an *En1*^{Cre} lacZ or green fluorescent protein (GFP) reporter system⁹, which also showed that significantly fewer En1-derived V1 neurons are present at E18.5 (Fig. 1b, f). Both *Pax6*^{−/−} and *En1-DTA* mice had normal numbers of lumbar spinal motor neurons that were appropriately organized into lateral and medial motor columns. (Fig. 1c, g). Spinal cords from *Pax6*^{−/−} mice did show an increase in commissural V0 and V3 interneuron cell numbers, coupled with a slight decrease in ipsilaterally projecting Chx10-positive V2 neurons¹³ (Chx10 is a marker of V2 neurons; Supplementary Fig. S1). However, there were no changes in ventral

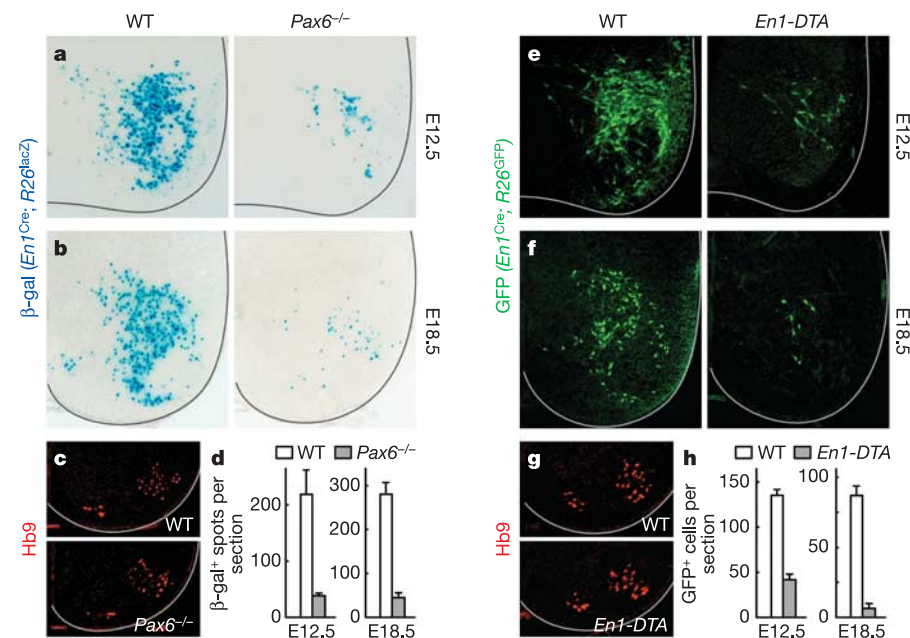


Figure 1 | V1 neurons are reduced in *Pax6*^{−/−} and *En1-DTA* spinal cords. **a, b**, Cross-sections through E12.5 (**a**) and E18.5 (**b**) wild-type (WT) and *Pax6*^{−/−} spinal cords, showing fewer En1-positive V1 neurons in the *Pax6*^{−/−} spinal cord. V1 neurons were marked using *En1*^{Cre}; *R26*^{lacZ} reporter alleles. **c**, Lumbar-level Hb9 staining of motor neurons is largely unaffected in *Pax6*^{−/−} embryos. **d**, Quantification of V1 neuron loss in the *Pax6*^{−/−} spinal cord. **e, f**, *En1-DTA* embryos show a selective loss of En1-positive V1 neurons at E12.5 (**e**) and E18.5 (**f**). **g**, Numbers of Hb9-positive motor neurons and columnar organization are normal in the *En1-DTA* spinal cord. **h**, Quantification of V1 neuron loss in the *En1-DTA* cord. Error bars indicate s.d.

¹Molecular Neurobiology Laboratory and ²System Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ³Mammalian Locomotor Laboratory, Department of Neuroscience, The Karolinska Institute, Retzius vag 8, Stockholm 17177, Sweden. ⁴Centre for Molecular Neurobiology, University of Hamburg, Hamburg 20251, Germany. [†]Present address: Department of Developmental Genetics, The Skirball Institute, New York University, New York, New York 10016, USA.

*These authors contributed equally to this work.

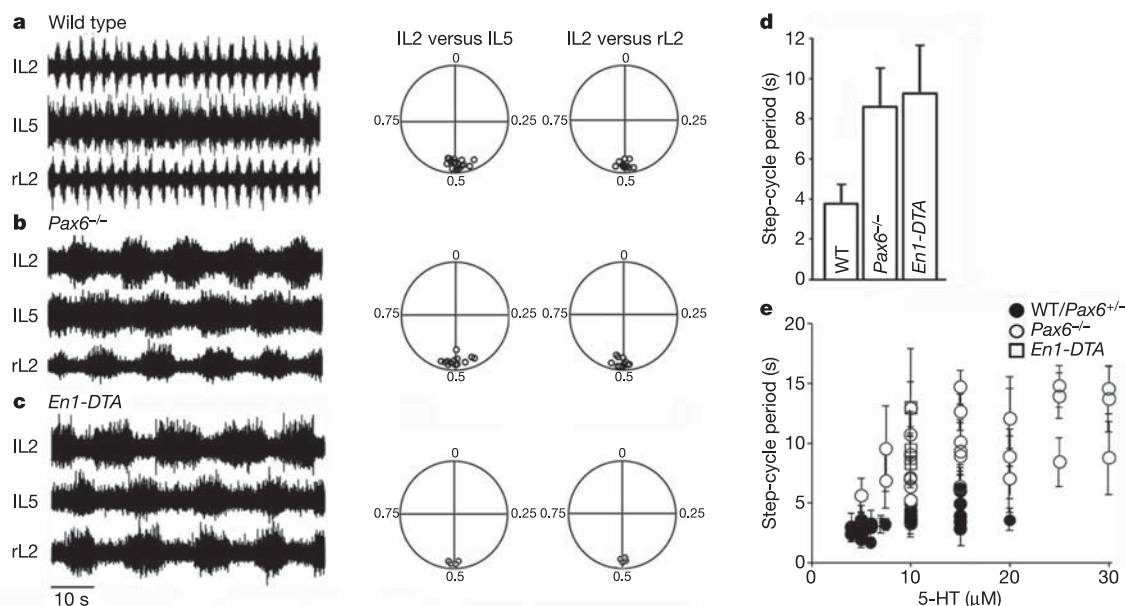


Figure 2 | Mice lacking V1 neurons show slow locomotor-like activity. **a–c**, Left, electrophysiological traces from left (IL2) and right (rL2) L2 and left L5 (IL5) ventral roots after application of 5 μ M NMDA and 10 μ M 5-HT, showing locomotor-like activity in E18.5 spinal cords from wild-type (**a**), *Pax6*^{−/−} (**b**) and *En1-DTA* (**c**) mice. Right, circular plots showing the

coupling of bursts in IL5 and rL2 with respect to IL2. Each point represents the vector point for one experiment. **d**, Mean step-cycle period in wild-type, *Pax6*^{−/−} and *En1-DTA* spinal cords. **e**, Plot of step-cycle period versus 5-HT concentration for *Pax6*^{−/−} (open circles), *En1-DTA* (open squares) and wild-type/*Pax6*^{+/−} spinal cords (filled circles). Error bars indicate s.d.

interneuron specification in the *En1-DTA* mice (Supplementary Fig. S1).

Limbic walking movements in vertebrates are characterized by the repetitive oscillatory bursting of motor neurons, in which flexor–extensor and left–right motor activities alternate. These locomotor-like oscillations can be induced in the isolated spinal cord by excitatory neurotransmitter agonists, and typically show a step-cycle period of 2–4 s (refs 5, 14, 15 and Fig. 2a). We asked whether V1 neurons regulate the coordination of flexor–extensor motor neurons during locomotion, because many V1 neurons differentiate into Renshaw cells and Ia inhibitory interneurons, two inhibitory cell types that have been proposed to coordinate flexor and extensor muscle activity around ipsilateral limb joints^{9,16–19}. Spinal cords lacking V1 neurons showed a normal pattern of alternating flexor (L2) activity and extensor (L5) activity (Fig. 2a–c). Left–right alternation (that is, IL2 versus rL2) was also normal. However, ‘locomoting’ *Pax6*^{−/−} and *En1-DTA* spinal cords did show a significant lengthening of both the step-cycle period and motor neuron burst duration (Fig. 2b–e).

Whereas 5 μ M NMDA (*N*-methyl-D-aspartate) in combination with 5 μ M serotonin (5-hydroxytryptamine or 5-HT) typically induces robust locomotor-like activity in wild-type spinal cords¹⁵ (Fig. 2a), we were only able to elicit a stable pattern of locomotor-like activity in *Pax6*^{−/−} spinal cords by increasing the concentration of 5-HT to 7.5–20 μ M (Fig. 2b, e). In experiments using 10 μ M 5-HT, the step-cycle period (8.4 ± 2.9 s, mean $\pm$ s.d.) and burst duration (4.4 ± 1.0 s) in *Pax6*^{−/−} spinal cords ($n = 7$) were increased significantly ($P < 0.001$, Student’s *t*-test) compared to wild-type and heterozygous (*Pax6*^{+/−}) spinal cords (step-cycle period 3.8 ± 0.8 s, burst duration 1.7 ± 0.8 s; $n = 10$). This relative increase in the step-cycle period was seen over a wide range of 5-HT concentrations (Fig. 2e). Similar lengthening of the step cycle (Fig. 2c, d) and burst duration was seen in *En1-DTA* mice (9.3 ± 2.3 s and 5.4 ± 1.1 s, respectively; $n = 5$ spinal cords; $P < 0.001$, *t*-test). V1 neurons are therefore dispensable for flexor–extensor coordination, but have an essential role in determining the speed of the locomotor rhythm.

Current-clamp recordings of motor neurons in wild-type and *Pax6*^{−/−} spinal cords (Fig. 3a–d) revealed that *Pax6*^{−/−} motor

neurons show prolonged periods of membrane potential depolarization, and that they continue to fire action potentials throughout this depolarized phase (Fig. 3c). This is consistent with flexor- and extensor-related CPG neurons remaining active for longer periods of time. These intracellular recordings also revealed that the alternating excitatory and inhibitory drive that normally underlies rhythmic changes in motor neuron membrane potential²⁰ is still present in the *Pax6*^{−/−} spinal cord (Fig. 3e). Importantly, *Pax6*^{−/−} motor neurons still receive the phasic inhibitory inputs that are essential for coordinating alternating flexor and extensor activity (Fig. 3f). The origin of these phasic inhibitory inputs is not known—although they could in principle be derived from contralateral commissural interneurons^{21,22}, this seems not to be the case, as ipsilateral locomotor coordination is normal in the *Pax6*^{−/−} cord after spinal cord hemisection (data not shown).

We then used *En1*^{tlz/tlz} null ‘knock-in’ mice (see ref. 8) to address the function of V1 neurons in the intact animal. Spinal cords isolated from E18.5 *En1*^{tlz/tlz} mice show a 3–4-fold slowing of the locomotor step cycle and lengthening of the burst duration (Fig. 4a, b; 11.2 ± 2.3 s and 6.1 ± 1.8 s, respectively; $n = 10$ spinal cords), which closely resemble the behavioural changes seen in *Pax6*^{−/−} and *En1-DTA* spinal cords (compare with Fig. 2). Using a *Wnt1*^{En1} transgene²³ to rescue the *En1* mutant midbrain–hindbrain phenotype, we obtained adult *En1*^{tlz/tlz} mice in which the V1 neuron connectivity defects in the spinal cords^{8,9} are not rescued (H.S., unpublished observation). When rotarod tests were performed on 4–6-month-old *En1*^{tlz/tlz}; *Wnt1*^{En1} mice and age-matched controls, *En1*^{tlz/tlz}; *Wnt1*^{En1} mice showed a clear deficit in their ability to walk and maintain their balance at higher rotarod speeds (Fig. 4c; see Methods). Although the duration (and the speed) for which the mutant animals walked on the rotarod increased during the trial period, at no time did they attain walking speeds above 18 r.p.m. However, these animals did walk on the rotarod for extended periods of times at idling speed (5 r.p.m., data not shown). In contrast, wild-type animals walked and maintained their balance at much higher speeds (approximately 22 r.p.m. on day 1 and up to 35 r.p.m. on day 5). This suggests that adult *En1* mutants are impaired in their ability to perform quick stepping movements.

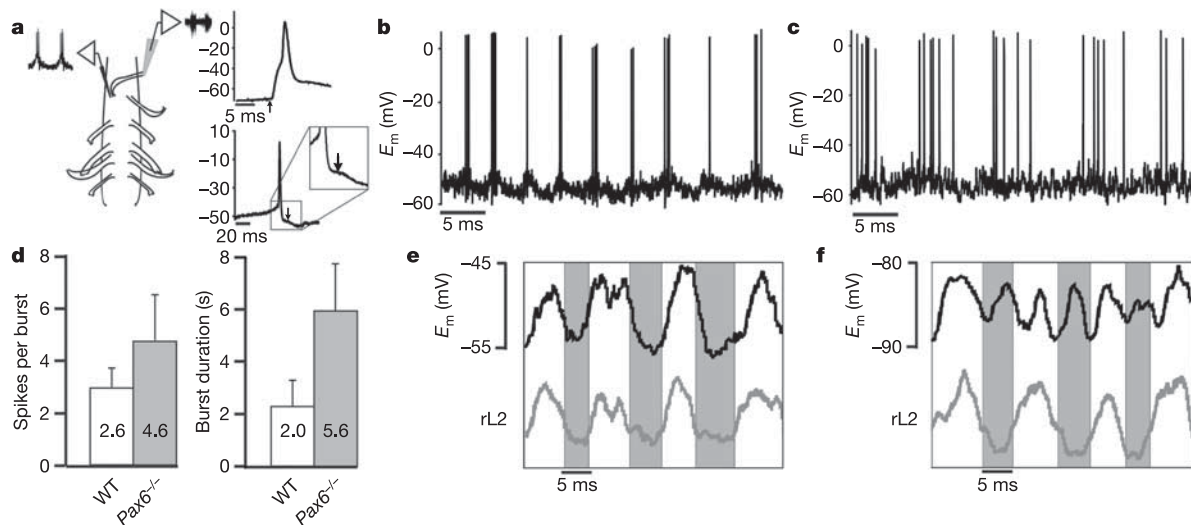


Figure 3 | Motor neuron firing during locomotion in the *Pax6*^{-/-} spinal cord is prolonged and shows rhythmic inhibition. **a**, Schematic of recording setup, with a suction electrode recording electroneurogram activity from rL2 ventral root and an intracellular electrode recording from a motor neuron located in L2. Motor neurons were identified by the presence of an antidromic action potential in response to ventral root stimulation (upper right panel, arrow) and by a small after-depolarization following orthodromic spikes (lower right panel, arrow), which is unique to rodent motor neurons. **b**, **c**, Recording from motor neurons located in rL2 of wild-type (**b**) and *Pax6*^{-/-} (**c**) mice after application of NMDA and 5-HT. **d**, Chart showing the mean number of action potentials (mean $\pm$ s.d.) fired per depolarization of the motor neuron and the mean duration of the

depolarized phase (LDP) of the motor neuron during locomotion in wild-type and *Pax6*^{-/-}. **e**, **f**, Intracellular recording from a motor neuron located in the rL2 segment (upper traces) and electroneurogram recording from the rL2 ventral root (lower traces, rectified and integrated) in a *Pax6*^{-/-} mouse. Grey area indicate the period during which rL2 is inactive. In **e**, recordings were performed at resting membrane potential (E_m -50 mV) after application of 5 μ M NMDA + 15 μ M 5-HT. The motor neuron is active in-phase with ventral root activity. In **f**, the motor neuron is hyperpolarized to -88 mV, beyond the calculated Cl^- reversal potential, which results in the emergence of out-of-phase depolarization owing to reversed inhibitory currents.

Next, we used transgenic mice (*AlstR192* mice) that conditionally express the *Drosophila* allatostatin G-protein-coupled receptor (*AlstR*) to test whether acutely suppressing V1 neuron excitability causes a similar lengthening of the locomotor step cycle (Fig. 5a). *AlstR* couples to endogenous mammalian GIRK channels (inward-rectifying K^+ channels), causing a decrease in cellular input resistance and neuronal excitability²⁴. In control experiments, allatostatin (100 nM–5 μ M) had no effect on motor rhythm generation, locomotor coordination or the step-cycle period (Supplementary Fig. S2a). In contrast, spinal cords from *nestin*^{Cre}; *AlstR192* mice showed strong allatostatin-dependent depression in rhythmic motor activity (Supplementary Fig. S2b).

En1^{Cre}; *AlstR192* mice were then used to selectively express *AlstR* in V1 neurons (Fig. 5a, b). In spinal cord slices prepared from *En1*^{Cre}; *AlstR192* mice, we observed a pattern of GFP reporter expression that was identical to the normal distribution of V1 neurons (Fig. 5c). Whole-cell recordings from identified GFP-positive V1 neurons ($n = 7$ cells) revealed a reversible decrease in neuronal excitability in response to current steps (Fig. 5d) and ramps (Fig. 5e) after the addition of allatostatin (10 nM). The decrease in excitability (2–3-fold), although not as pronounced as in isolated adult neurons²⁴, was nevertheless substantial, and it was not seen in control, *AlstR*-negative neurons (Supplementary Fig. S2c).

When locomotor-like activity was induced in postnatal day P0–P2 *En1*^{Cre}; *AlstR192* spinal cords (5 μ M NMDA + 10 μ M 5-HT; Fig. 5f), we observed a step-cycle period of 3–4 s (3.3 ± 0.8 s, $n = 15$ spinal cords), comparable to that seen in wild-type animals of a similar age. Application of allatostatin (1–5 μ M) to these spinal cords resulted in marked and significant lengthening of the step-cycle period (7.9 ± 1.8 s, $n = 15$ spinal cords; $P = 0.002$, paired *t*-test), which was reversed upon washing out the neuropeptide (Fig. 5f). As such, lengthening of the locomotor step cycle that results from the defects in V1 neuron development can be reproduced when V1 neuronal activity is suppressed. Two recent studies have provided evidence that modifying neuronal activity during development can alter the

neurotransmitter phenotype of spinal neurons²⁵ and produce small changes in CPG motor outputs²⁶, indicating that activity may have a role in organizing these locomotor networks. However, our results show that the changes in locomotor CPG outputs following the acute

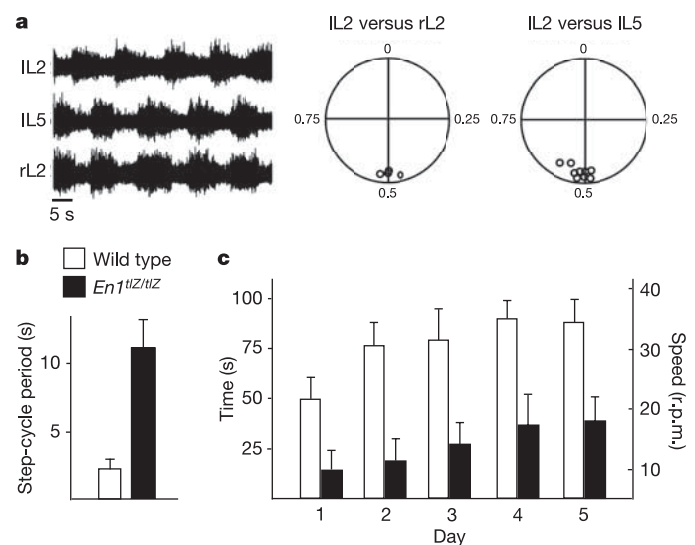


Figure 4 | *En1*^{tlz/tlz} mice show a slowing of stepping movements. **a**, Left, electroneurogram recordings from IL2, IL5 and rL2 ventral roots after application of 5 μ M NMDA/7.5 μ M 5-HT to an isolated E18.5 *En1*^{tlz/tlz} spinal cord. Right, circular plots indicate normal ventral root alternation. Each small circle corresponds to the vector point for a single experiment. **b**, Graph showing the increase in the step-cycle period in *En1*^{tlz/tlz} mutants compared with wild-type spinal cords. **c**, Rotarod testing of mouse motor behaviours, showing the time and bar speed that age-matched wild-type (open bars) and *En1*^{tlz/tlz}; *Wnt1*^{En1} mice (filled bars; three animals each) were able to run on an accelerating rotarod. Error bars in **b**, **c** show s.d.

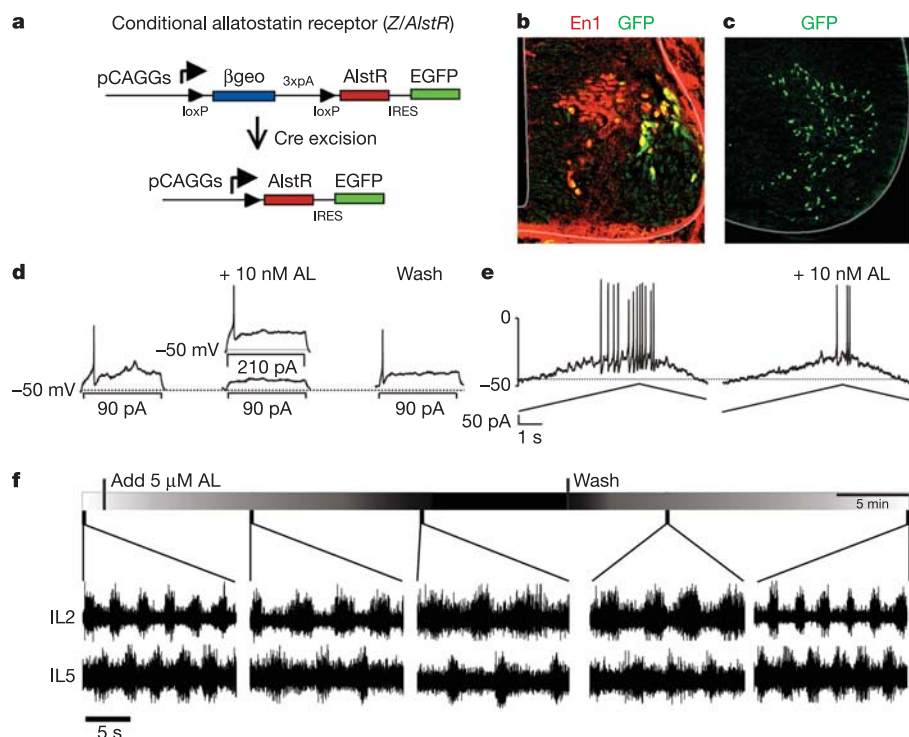


Figure 5 | Acute silencing of V1 neurons via the allatostatin receptor causes a slowing of locomotor-like rhythmicity. **a**, Schematic showing the allatostatin receptor (*Z/AlstR*) transgene. Crossing mice that harbour this transgene with an *En1*^{Cre} mouse excises the loxP-flanked β -geo sequences, leading to transcription of *AlstR* in *En1*-derived cells. **b**, Co-expression of GFP (green) and *En1* (red) in an E11.5 *En1*^{Cre}; *AlstR192* mouse, showing V1 cells expressing *AlstR*. Many newborn V1 neurons in the subventricular zone are *AlstR*/GFP-negative owing to a delay in *En1*-dependent Cre removal of the stop cassette. **c**, Spinal cord slice from a P0 *En1*^{Cre}; *AlstR192* mouse carrying the *ZnG* reporter allele, showing GFP expression in V1

neurons. **d**, Recording from a V1 neuron expressing *AlstR* in a P1 *En1*^{Cre}; *AlstR192*; *ZnG* mouse in response to current steps. **e**, Recordings from an *AlstR*-expressing V1 neuron in response to a current ramp. Before allatostatin (AL) application, the cell begins to fire upon current injection of 105 pA. After allatostatin application (10 nM), the cell only fires when the injected current reaches 190 pA. **f**, Electroneurogram recordings from IL2 and IL5 ventral roots in a P1 *En1*^{Cre}; *AlstR192* mouse, showing the effects of allatostatin (5 μ M) application and wash-out on locomotor activity in the isolated spinal cord. The lengthening of the step cycle after allatostatin application was significant ($P = 0.02$, paired *t*-test; $n = 15$ spinal cords).

suppression of V1 neurons compared with the early loss of V1 neurons are highly concordant. This suggests that the absence of V1 neurons during development does not markedly reconfigure the spinal locomotor CPG network.

Our findings outline an essential role for V1 neurons in mammals in regulating the duration of the locomotor step cycle, and hence the speed of the locomotion. This function is specific for V1 neurons, as we have not seen marked changes in the frequency or duration of the step cycle when either V0 or V3 neurons are deleted from the spinal cord⁵ (M.G., unpublished observations). Our conclusion that V1 neurons are necessary for 'fast' locomotor outputs in mice is supported by studies in *Xenopus* tadpoles showing that aIN neurons, the homologues of mouse V1 neurons, are the primary source of early cycle inhibition to the swimming locomotor CPG¹⁰. Although the role of aIN neurons in regulating the speed of *Xenopus* swimming movements was not examined, a strong correlation was seen between aIN-derived inhibitory inputs to CPG neurons and the frequency of swimming movements¹⁰. This raises the possibility that the *Xenopus* homologues of V1 neurons may facilitate fast swimming movements, in much the same way that their mammalian counterparts are required for fast 'walking' movements.

The exact contribution individual V1 interneuron subtypes make to regulating the duration of motor neuron bursting during fictive locomotion remains to be determined. Inhibition of Renshaw cells with cholinergic blockers causes a small increase in the step-cycle period²⁶ (S.G. and O.K., unpublished observations), suggesting that a group of non-Renshaw V1 neurons or the V1 population as a whole regulates the step-cycle period during fictive locomotion. Although Renshaw cells and Ia inhibitory interneurons have described roles in

coordinating flexor–extensor motor neurons during spinal reflexes²⁷, our study demonstrates that the V1 population as a whole does not have a primary role in regulating locomotor CPG flexor–extensor activity. It therefore seems that an additional group of non-V1 interneurons have been incorporated into the walking CPG of terrestrial vertebrates to secure flexor–extensor alternation.

Our study also demonstrates that the acute silencing of a select population of neurons using genetic approaches can be used to elucidate their function with respect to a defined behaviour such as locomotion. Our results highlight the feasibility of using ligand-activated GIRK channels to selectively manipulate neuronal excitability in the vertebrate nervous system, and the conditional *AlstR192* mice described here can be used to selectively silence neurons throughout the nervous system. This and similar genetic approaches now make it feasible to selectively probe the function of small populations of neurons, which should facilitate the mapping of neural circuits at higher resolutions than was previously possible.

METHODS

Animals. The generation and genotyping of the *Pax6*^{lacZ}, *Sey*, *En1*^{Cre}, *En1*^{tlz}, *Wnt1*^{En1}, *nestin*^{Cre}, *R26-lacZ*^{flox/DTA}, *R26*^{lacZ} and *R26*^{GFP} alleles in mice has been previously described^{8,9,12,23,28}. Embryos and tissues were obtained from timed matings.

Generation of conditional *AlstR* mice. The *Drosophila AlstR* coding sequence, followed by an IRES-EGFP²⁴, was inserted downstream of a loxP-flanked β -geo/polyA stop sequence in the conditional *Z/EG* construct²⁹ to produce the *Z/AlstR* expression vector (see Fig. 4a). The linearized *Z/AlstR* construct was electroporated into the embryonic stem cell line W9.5. G418-resistant clones were screened for uniform and robust lacZ expression, and for single transgene integrants. Two clones gave germline transmission (*AlstR172* and *AlstR192*).

AlstR192 mice that showed lacZ expression throughout the nervous system were used for all further experiments.

In situ hybridization and immunohistochemistry. Immunostainings were performed as previously described⁵. Images were captured using a Zeiss LSM510 confocal microscope and assembled using Adobe Photoshop. Cell numbers are indicated as mean $\pm$ s.d. per section per spinal cord side, and were determined on counts of at least three embryos (five sections each). *In situ* hybridization and staining for β -galactosidase activity were performed as previously described^{5,8}.

Electrophysiology: electroneurogram recordings. Electrophysiological experiments were performed on E18.5–P2 mice in accordance with the ethical rules stipulated by NIH and the Swedish government. Animals were anaesthetized and decapitated, and spinal cords were dissected out in ice-cold Ringers' solution⁴. Recordings were made in Ringers' solution at 20 °C by placing bipolar suction electrodes on three of the second and fifth lumbar ventral roots (that is rL2, IL2, rL5, IL5). Electroneurogram signals were amplified, band-pass filtered (100 Hz–1 kHz), digitized and collected using Axoscope software (Axon Instruments). Rhythmic locomotor activity was induced by adding NMDA (5 μ M) and 5-HT (5–30 μ M) to the Ringers' solution. Effects of allatostatin (10 nM–5 μ M) were examined by adding the peptide to the perfusion solution.

Electrophysiology: analysis of locomotor activity. Step-cycle period and burst duration were determined by analysing IL2 or rL2 activity⁴. Step-cycle period and burst duration averages were determined from all recorded bursts after the onset of stable locomotor-like activity. The effects of allatostatin were measured 10 min after application to allow for drug wash-in. All measurements are given $\pm$ s.d. *P* values less than 0.05 were considered significant. Circular statistics¹⁹ were used to determine the coupling strength between left and right ventral roots, as well as between flexor-related (L2) and extensor-related (L5) ventral roots on the same side of the spinal cord. L2 bursts were used as a reference. Vector points representing the phase values between 0 and 1 were plotted for each experiment, and show the mean phase as well as the concentration of phase values around the mean. Appropriate left–right or flexor–extensor alternation is represented by phase values around 0.5.

Electrophysiology: intracellular recordings. Whole-cell patch-clamp recordings from motor neurons were made using recording pipettes with a resistance of 4–5 M Ω . The microelectrode was driven into the spinal cord through a pial patch in the ventrolateral surface. Cells were patched in the motor neuron area and recorded using an AxoClamp 2B amplifier (Axon Instruments).

Electrophysiology: acute spinal cord slice recordings. For slice recordings, spinal cord slices 250–300- μ m thick were cut using a Leica VY1000E vibrating microtome. After a 1-h recovery period, slices were transferred to a recording chamber, mounted on an Olympus BX51W1 microscope and perfused with oxygenated Ringers' solution at room temperature. GFP-positive cells were visualized using a DAGE-MTI IR-1000 CCD camera and patched visually using a Sutter MPC-325 micromanipulator. Recordings were made in current-clamp mode using a MultiClamp 700B amplifier (Axon Instruments).

Behavioural testing. Rotarod test were performed during the light phase of a 12 h:12 h light:dark cycle as previously described³⁰. Mice were placed on a Rotamex 4/8 (Columbus Instruments) idling at 5 r.p.m. Rotarod speed was set to increase gradually from 5 to 65 r.p.m. over the course of 3 min. Retention time on the rod was recorded for three trials per mouse per day for five consecutive days, using three mutant and three wild-type mice.

Received 15 November; accepted 22 December 2005.

- Grillner, S. The motor infrastructure: from ion channels to neuronal networks. *Nature Rev. Neurosci.* **4**, 573–586 (2003).
- Kiehn, O. & Kjaerulff, O. Distribution of central pattern generators for rhythmic motor outputs in the spinal cord of limbed vertebrates. *Ann. NY Acad. Sci.* **860**, 110–129 (1998).
- Roberts, A., Soffe, S. R., Wolf, E. S., Yoshida, M. & Zhao, F. Y. Central circuits controlling locomotion in young frog tadpoles. *Ann. NY Acad. Sci.* **860**, 19–34 (1998).
- Butt, S. J., Harris-Warrick, R. M. & Kiehn, O. Firing properties of identified interneuron populations in the mammalian hindlimb central pattern generator. *J. Neurosci.* **22**, 9961–9971 (2002).
- Lanuza, G. M., Gosgnach, S., Pierani, A., Jessell, T. M. & Goulding, M. Genetic identification of spinal interneurons that coordinate left-right locomotor activity necessary for walking movements. *Neuron* **42**, 375–386 (2004).
- Angel, M. J., Jankowska, E. & McCreary, D. A. Candidate interneurons mediating group I disynaptic EPSPs in extensor motoneurons during fictive locomotion in the cat. *J. Physiol. (Lond.)* **563**, 597–610 (2005).
- Jessell, T. M. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Rev. Genet.* **1**, 20–29 (2000).

- Saueressig, H., Burrill, J. & Goulding, M. Engrailed-1 and netrin-1 regulate axon pathfinding by association interneurons that project to motor neurons. *Development* **126**, 4201–4212 (1999).
- Sapir, T. *et al.* *Pax6* and *Engrailed-1* regulate two distinct aspects of Renshaw cell development. *J. Neurosci.* **24**, 1255–1264 (2004).
- Li, W. C., Higashijima, S., Parry, D. M., Roberts, A. & Soffe, S. R. Primitive roles for inhibitory interneurons in developing frog spinal cord. *J. Neurosci.* **24**, 5840–5848 (2004).
- Higashijima, S., Masino, M. A., Mandel, G. & Fetcho, J. R. Engrailed-1 expression marks a primitive class of inhibitory spinal interneuron. *J. Neurosci.* **24**, 5827–5839 (2004).
- Brockschneider, D. *et al.* Cell depletion due to diphtheria toxin fragment A after Cre-mediated recombination. *Mol. Cell. Biol.* **24**, 7636–7642 (2004).
- Ericson, J. *et al.* *Pax6* controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**, 169–180 (1997).
- Jiang, Z., Carlin, K. P. & Brownstone, R. M. An *in vitro* functionally mature mouse spinal cord preparation for the study of spinal motor networks. *Brain Res.* **816**, 493–499 (1999).
- Kullander, K. *et al.* Role of EphA4 and EphrinB3 in local neuronal circuits that control walking. *Science* **299**, 1889–1892 (2003).
- Wenner, P., O'Donovan, M. J. & Matisse, M. P. Topographical and physiological characterization of interneurons that express engrailed-1 in the embryonic chick spinal cord. *J. Neurophysiol.* **84**, 2651–2657 (2000).
- Alvarez, F. J. *et al.* Postnatal phenotype and localization of V1-derived interneurons. *J. Comp. Neurol.* **493**, 177–192 (2005).
- Jankowska, E., Fu, T. C. & Lundberg, A. Reciprocal Ia inhibition during the late reflexes evoked from the flexor reflex afferents after DOPA. *Brain Res.* **85**, 99–102 (1975).
- Kjaerulff, O. & Kiehn, O. Distribution of networks generating and coordinating locomotor activity in the neonatal rat spinal cord *in vitro*: a lesion study. *J. Neurosci.* **16**, 5777–5794 (1996).
- Kiehn, O., Hounsgaard, J. & Sillar, K. T. in *Neurons, Networks, and Motor Behaviour* (eds Stein, P. S. G., Stuart, D., Selverston, A. & Grillner, S.) 47–59 (MIT Press, Cambridge, Massachusetts, 1997).
- Kjaerulff, O. & Kiehn, O. Crossed rhythmic synaptic input to motoneurons during selective activation of the contralateral spinal locomotor network. *J. Neurosci.* **17**, 9433–9447 (1997).
- Butt, S. J. & Kiehn, O. Functional identification of interneurons responsible for left-right coordination of hindlimbs in mammals. *Neuron* **38**, 953–963 (2003).
- Danielian, P. S. & McMahon, A. P. *Engrailed-1* as a target of the Wnt-1 signalling pathway in vertebrate midbrain development. *Nature* **383**, 332–334 (1996).
- Lechner, H. A., Lein, E. S. & Callaway, E. M. A genetic method for selective and quickly reversible silencing of mammalian neurons. *J. Neurosci.* **22**, 5287–5290 (2002).
- Borodinsky, L. N. *et al.* Activity-dependent homeostatic specification of transmitter expression in embryonic neurons. *Nature* **429**, 523–530 (2004).
- Myers, C. P. *et al.* Cholinergic input is required during embryonic development to mediate proper assembly of spinal locomotor circuits. *Neuron* **46**, 37–49 (2005).
- Burke, R. E. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 77–120 (Oxford Univ. Press, Oxford, 1998).
- St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A. & Gruss, P. *Pax6* is required for differentiation of glucagon-producing α -cells in mouse pancreas. *Nature* **387**, 406–409 (1997).
- Novak, A., Guo, C., Yang, W., Nagy, A. & Lobe, C. G. Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon cre-mediated excision. *Genesis* **28**, 147–155 (2000).
- Brandon, E. P. *et al.* Choline transporter 1 maintains cholinergic function in choline acetyltransferase haploinsufficiency. *J. Neurosci.* **24**, 5459–5466 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Rivier for the allatostatin peptide, P. Gruss and A. Mansouri for *Pax6* mutant mice, P. Soriano for *Rosa26^{lacZ}* mice, A. McMahon for *Wnt1^{En1}* mice, S. Narayan for generating ZNG reporter mice, and M. Gross, A. Joyner, S. Pfaff and P. Slesinger for materials. We thank G. Lemke, S. Pfaff, P. Gray and K. Quinlan for their comments on the manuscript. This research was supported by grants from the National Institutes of Health (M.G., O.K.) and the Human Frontiers Science Program (O.K. and M.G.). G.M.L. was supported by an HFSP postdoctoral fellowship.

Author Contributions are provided in the Supplementary Information.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.G. (goulding@salk.edu).

LETTERS

CLC-7 requires Ostm1 as a β -subunit to support bone resorption and lysosomal function

Philipp F. Lange¹, Lena Wartosch¹, Thomas J. Jentsch¹ & Jens C. Fuhrmann¹

Mutations in CLC-7, a late endosomal/lysosomal member of the CLC family of chloride channels and transporters^{1,2}, cause osteopetrosis³ and lysosomal storage disease⁴ in humans and mice. Severe osteopetrosis is also observed with mutations in the *OSTM1* gene, which encodes a membrane protein of unknown function⁵. Here we show that both CLC-7 and Ostm1 proteins co-localize in late endosomes and lysosomes of various tissues, as well as in the ruffled border of bone-resorbing osteoclasts. Co-immunoprecipitations show that CLC-7 and Ostm1 form a molecular complex and suggest that Ostm1 is a β -subunit of CLC-7. CLC-7 is required for Ostm1 to reach lysosomes, where the highly glycosylated Ostm1 luminal domain is cleaved. Protein but not RNA levels of CLC-7 are greatly reduced in grey-lethal mice, which lack Ostm1, suggesting that the CLC-7–Ostm1 interaction is important for protein stability. As CLC-7 protein levels in Ostm1-deficient tissues and cells, including osteoclasts, are decreased below 10% of normal levels, *Ostm1* mutations probably cause

osteopetrosis by impairing the acidification of the osteoclast resorption lacuna, which depends on CLC-7 (ref. 3). The finding that grey-lethal mice, just like CLC-7-deficient mice⁴, show lysosomal storage and neurodegeneration in addition to osteopetrosis implies a more general importance for CLC-7–Ostm1 complexes.

In western blots of brain membranes, an antibody against the carboxy terminus of mouse Ostm1 recognized an ~80 kDa band and a doublet at ~35–45 kDa (Fig. 1c). These bands were absent in grey-lethal (*gl*) mice, an osteopetrotic mouse mutant⁶ carrying a deletion of the *Ostm1* promoter and exon 1 (ref. 5). Incubating cells with protease inhibitors increased the proportion of the large band at the expense of the smaller ones (Fig. 1d), indicating that the small forms of Ostm1 are produced by proteolytic cleavage of the ~80 kDa protein. The apparent sizes of the Ostm1 species suggest cleavage roughly in the middle of the protein (Fig. 1b). As the large and the small forms (Fig. 1c) ran together in a single large band under non-reducing conditions (Fig. 1e), the cleaved fragments might be linked

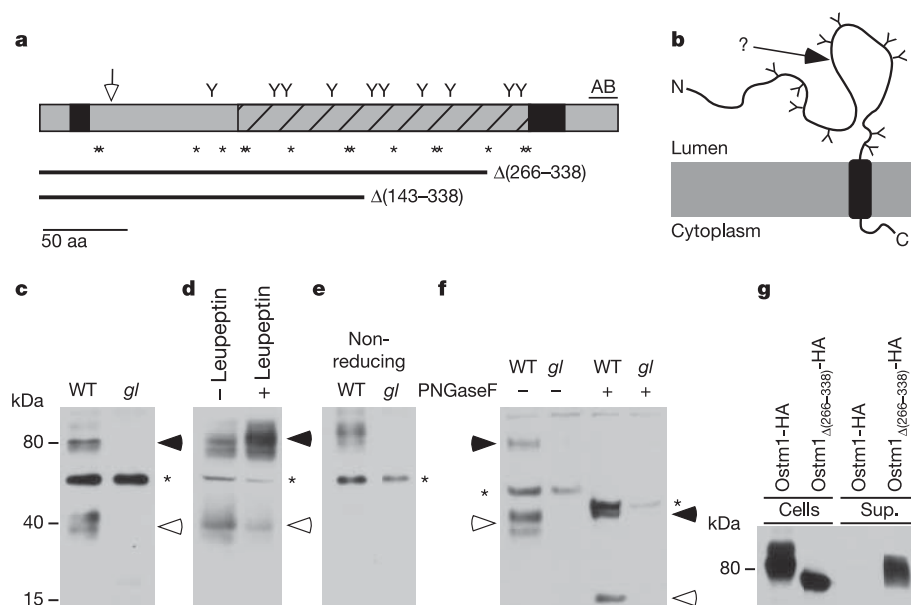


Figure 1 | Structural features of the Ostm1 protein. **a**, Schematic of Ostm1 protein. Black box, hydrophobic stretch; dashed box, proposed RING-finger domain⁷; Y, consensus site for N-linked glycosylation. Asterisks indicate cysteine residues; arrow indicates predicted signal peptide cleavage site; AB indicates antibody binding site. Lines below show the truncated proteins predicted from human *OSTM1* mutations^{5,20}. **b**, Topology model derived from our work. Arrow shows the approximate cleavage site in lysosomal Ostm1. **c–f**, Western blots of Ostm1. WT, wild type. **c**, The band at ~80 kDa (filled arrowhead) and the doublet at ~35–45 kDa detected in wild-type brain membranes (open arrowhead) were absent from the grey-lethal brain.

Asterisk indicates a non-specific band. **d**, Incubating fibroblasts with the protease inhibitor leupeptin increased the proportion of the large Ostm1 species. Similar results were obtained with E64, which inhibits several cathepsins (not shown). **e**, Western blots of brain proteins separated by non-reducing SDS–PAGE showed a single band of ~80 kDa Ostm1. **f**, Deglycosylation with PNGaseF reduced the sizes of all Ostm1 species detected under reducing conditions. **g**, Western blot of cells (lanes 1 and 2) and supernatants (sup., lanes 3 and 4) from HEK293 cells expressing Ostm1 (lanes 1 and 3) or a truncated form of Ostm1 (lanes 2 and 4) that mimics a human mutation⁵. Both proteins carried a C-terminal HA-epitope.

¹Zentrum für Molekulare Neurobiologie Hamburg, ZMNH, Universität Hamburg, Falkenried 94, D-20246 Hamburg, Germany.

by disulphide bonds between some of the cysteine residues that abound in the luminal domain of Ostm1 (Fig. 1a).

Recent work has proposed an E3 ubiquitin ligase function for Ostm1 (cloned independently from rat as GIPN)⁷. The stretch between the amino- and carboxy-terminal hydrophobic regions of Ostm1 (Fig. 1a) shows weak homology to RING-finger proteins and was suggested to be cytosolic⁷. However, this stretch (and no other part of the protein) contains several consensus sites for N-linked glycosylation. Several or all of these sites are used, because deglycosylation with PNGaseF (peptide *N*-glycosidase F) greatly reduced the apparent size of all Ostm1 species (Fig. 1f). The observed glycosylation places the hypothetical RING-finger domain⁷ in the lumen of the endoplasmic reticulum (ER), a localization difficult to reconcile with the cytosolic/nucleoplasmic activity of ubiquitin ligases⁸. The disappearance in transfected cells of a haemagglutinin (HA)-epitope added to the N terminus indicates the presence of a cleavable signal peptide (data not shown). We also investigated an Ostm1 mutant truncated before the second hydrophobic stretch. This mutant, but not wild-type Ostm1, was secreted into the supernatant of transfected cells (Fig. 1g). Hence, the first and second hydrophobic domains serve as a cleavable signal peptide and transmembrane domain, respectively, in agreement with the type I transmembrane protein model proposed in ref. 5 (Fig. 1b). Figure 1f also showed that the apparent molecular mass of the largest deglycosylated band

agreed roughly with the prediction from the *Ostm1* reading frame (~37 kDa). As deglycosylation of the small Ostm1 species yielded a single band, the doublet is attributable to non-uniform glycosylation.

In cultured primary fibroblasts, Ostm1 co-localized with Lamp-1, a marker for late endosomes and lysosomes (Fig. 2a). This localization resembled that of CIC-7, the loss of which also causes osteopetrosis³. However, when fibroblasts were transiently transfected with Ostm1 (Fig. 2b), it showed an ER-like distribution and no significant co-localization with Lamp-1 was observed. Co-transfection with CIC-7 restored a punctate Ostm1 staining pattern that largely co-stained with Lamp-1 (Fig. 2c) and CIC-7 (Fig. 2d). The effect of CIC-7 was specific, as co-transfection with neither CIC-3 (not shown) nor CIC-6 (Supplementary Fig. S1), which are both expressed in the endosomal/lysosomal pathway^{2,9}, resulted in such changes.

As CIC-7 changed Ostm1 localization, we studied Ostm1 in CIC-7-knockout (*Clcn7*^{-/-}) mice. Western blots indicated exclusive loss of the 35–45 kDa Ostm1 doublet in brain tissue from *Clcn7*^{-/-} mice (Fig. 3a). Subcellular fractionation of wild-type brain tissue revealed that the small Ostm1 form was co-enriched with lysosomal markers (bottom fractions 1–2), whereas the 80 kDa form was only detectable in fractions 9–12, which contain markers for endosomes and the ER (Fig. 3b). Such experiments yielded samples containing only the small (wild-type fractions containing lysosomes) or large (*Clcn7*^{-/-}

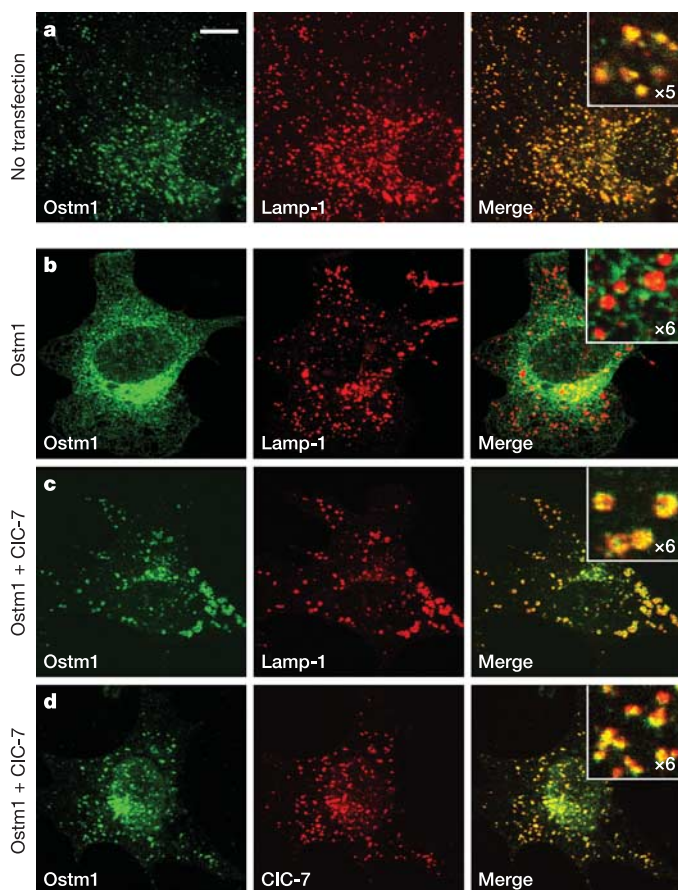


Figure 2 | CIC-7 is required for Ostm1 trafficking to lysosomes. **a**, Mouse fibroblast stained for Ostm1 and Lamp-1. Right panels show the overlay, insets show higher magnification. **b**, Ostm1-transfected fibroblasts showed reticular and perinuclear Ostm1 staining that co-localized poorly with Lamp-1. **c**, **d**, In fibroblasts co-transfected with Ostm1 and CIC-7, Ostm1 co-localized with Lamp-1 (**c**) and CIC-7 (**d**). In **b–d**, *Clcn7*^{-/-} fibroblasts were used to avoid effects of endogenous CIC-7, but similar results were seen in HeLa cells. Scale bar in **a** represents 8.5 μ m (**a**), 10 μ m (**b–d**), 1.7 μ m (insets).

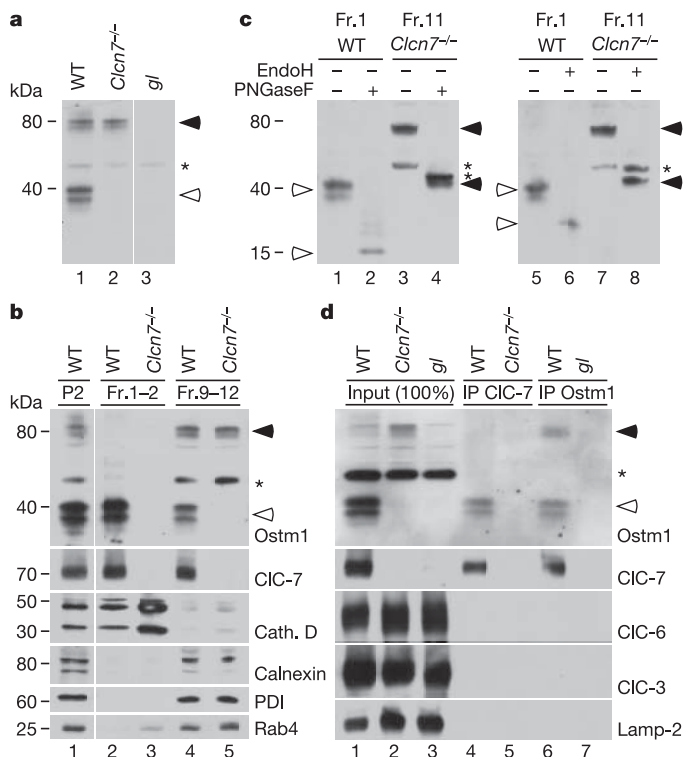


Figure 3 | CIC-7 and Ostm1 interact. **a**, Western blot of Ostm1 in brain from wild-type, *Clcn7*^{-/-} and *gl* mice. **b**, Subcellular distribution of Ostm1 in wild-type and *Clcn7*^{-/-} mice. Brain membranes (P2) and Percoll gradient fractions were analysed. Bottom fractions (fr. 1–2) contained lysosomal markers (cathepsin D, CIC-7), top fractions (9–12) contained ER (calnexin, PDI) and endosomal (Rab4) markers. Small forms of Ostm1 were enriched in lysosomal fractions, and the large form in ER/endosomal fractions. **c**, Deglycosylation of wild-type lysosomal and *Clcn7*^{-/-} ER/endosomal fractions, using EndoH or PNGaseF. **d**, Co-immunoprecipitation reveals a CIC-7–Ostm1 complex. Solubilized brain membranes were directly loaded on the gel (input, lanes 1–3), or first immunoprecipitated (IP) with CIC-7 (lanes 4, 5) or Ostm1 antibodies (lanes 6, 7). Western blots were probed for the proteins indicated on the right. Arrowheads show specific Ostm1 bands; asterisk shows non-specific bands. Equivalent amounts of lysates and precipitates were loaded.

fractions containing endosomes/ER) forms of Ostm1, which we used for deglycosylation experiments. PNGaseF reduced the size of both species (Fig. 3c, lanes 2 and 4), and only the 35–45 kDa form was partially resistant to EndoH (endoglycosidase H) (Fig. 3c, compare lanes 2 and 6). This indicates that the small, but not the large species of Ostm1 had left the ER. Our results thus suggest that CIC-7 is required for trafficking Ostm1 from the ER to lysosomes, and that the luminal domain of Ostm1 is cleaved on its way to, or in, this final compartment.

CIC-7 was efficiently co-immunoprecipitated from the brain with Ostm1, and vice versa (Fig. 3d). This interaction was specific, as it was observed neither with the related endosomal proteins CIC-3 and CIC-6, nor with Lamp-2. As expected from the lysosomal localization of CIC-7 (ref. 4), antibodies against CIC-7 almost exclusively precipitated the cleaved, lysosomal Ostm1 fragment from brain. Co-immunoprecipitation performed with transfected cells in which only the large form of Ostm1 could be detected showed that this putative ER form also interacted with CIC-7 (Supplementary Fig. S2). Notably, Fig. 3d also shows that CIC-7 levels are greatly reduced in brain extracts from *gl* mice (lane 3).

Both CIC-7 and Ostm1 are expressed in several tissues, including brain, liver, kidney and osteoclasts^{3,5,7,10} (see Supplementary Fig. S3). Immunohistochemistry of brain sections revealed that both proteins co-localize in neuronal cell bodies in structures that most likely represent lysosomes⁴ (Fig. 4a). Both proteins also co-localize in osteoclasts in a pattern that represents the 'ruffled border' (Fig. 4c). This acid-secreting plasma membrane domain was identified by co-staining for the $\alpha 3$ proton pump subunit^{11,12} (Fig. 4d), mutations of which also underlie osteopetrosis^{11,13–15}. Consistent with our western blot analyses (Fig. 3a, b, d), Ostm1 staining was very weak in *Cln7*^{-/-} mice (Supplementary Fig. S4), and CIC-7 labelling was greatly

reduced in grey-lethal cells. These cells included neurons (Fig. 4b) and osteoclasts (Fig. 4e), in which the remaining CIC-7 still localized to the ruffled border. Western blot analysis of brain, kidney, liver and bone revealed that CIC-7 levels were reduced to less than 10% in *gl* compared to wild-type mice (Fig. 4f, g). The transcript levels of both *Cln7* and *Ostm1* genes were unchanged (Supplementary Table).

CIC-7 might support bone resorption by electrically shunting the H^+ -ATPase that acidifies the osteoclast resorption lacuna³, or similarly by facilitating the insertion of proton-pump-containing vesicles into the ruffled border, which is underdeveloped in both *Cln7*^{-/-} and *gl* osteoclasts^{3,6}. This mechanism would also be feasible if CIC-7 were not a Cl^- channel, as believed so far^{3,10}, but an electrogenic Cl^-/H^+ -exchanger such as CIC-ec1 (ref. 16), CIC-4 or CIC-5 (refs 17, 18). As the intracellular localization of CIC-7 precluded biophysical analysis, these alternatives remain untested. Our work suggests that loss of OSTM1 causes osteopetrosis^{5,19,20} by decreasing the amount of CIC-7 to pathogenic levels. A 75% decrease in CIC-7 function as a result of dominant-negative mutations causes mild osteopetrosis in humans^{21,22}. The even lower levels of CIC-7 in *gl* mice may be sufficient to cause the severe osteopetrosis observed with *OSTM1* mutations^{5,19,20}.

Known disease-causing mutations in the human *OSTM1* gene^{5,20} introduce frame-shifts that replace the OSTM1 polypeptide with short unrelated sequences 143 or 21 residues^{5,20} before the transmembrane domain. When an epitope-tagged construct modelled after the latter mutation⁵ (Fig. 1a) was transfected into cells, co-expression with CIC-7 failed to direct the truncated Ostm1 to lysosomes (data not shown) and the truncated protein was secreted into the supernatant (Fig. 1g). We suggest that this mutant may lead to disease because Ostm1 no longer interacts with CIC-7.

The phenotypes of *Cln7*^{-/-} and *gl* mice are strikingly similar. On

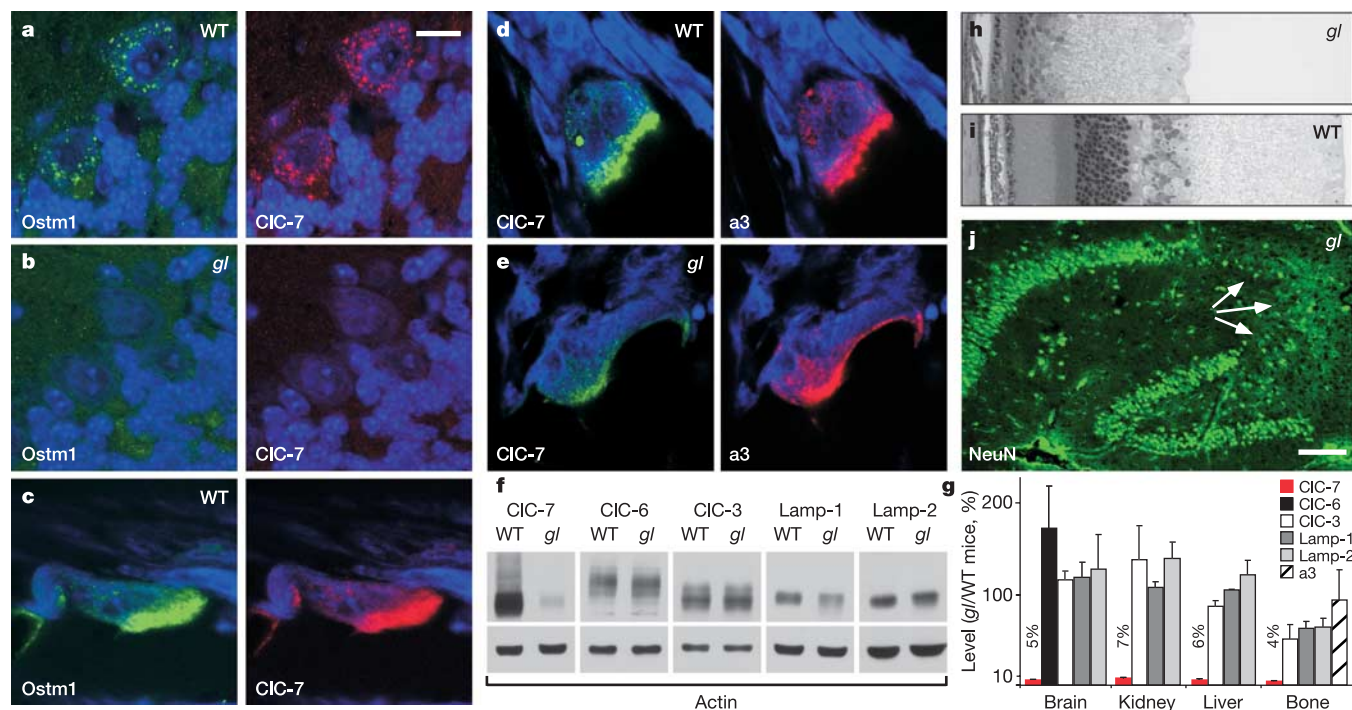


Figure 4 | CIC-7 and Ostm1 co-localization *in vivo* and *Cln7*^{-/-}-like phenotype of *gl* mice. **a, b, Immunofluorescence of cerebellar Purkinje cells. **a**, Ostm1 co-localized with CIC-7 in late endosomes/lysosomes⁴ of wild-type cells. **b**, In *gl* neurons, both proteins were undetectable. **c–e**, Immunofluorescence of osteoclasts *in situ*. Ostm1 and CIC-7 co-localized in the ruffled border (**c**), as did CIC-7 and the $\alpha 3$ proton-pump subunit (**d**). **e**, In *gl* osteoclasts, CIC-7 levels were greatly reduced. In **a–e**, nucleic acids were counterstained using TOTO (blue). Scale bar, 10 μm (**a–e**). **f**, Western blots show a decrease in CIC-7, but not in CIC-3, CIC-6,**

Lamp-1 or Lamp-2 in *gl* brain. **g**, Quantification of western blots from various tissues shows a decrease in CIC-7 in *gl* mice at postnatal day (P) 11–33, but no decrease in other late endosomal/lysosomal proteins. The moderate decrease in control proteins in *gl* bone might be explained by the osteopetrosis. Error bars indicate s.e.m.; $n = 3–9$ mouse pairs, except for CIC-3 in bone ($n = 2$). **h, i**, Retinal degeneration is visible in *gl* (**h**) but not in wild-type (**i**) mice at P31. **j**, NeuN (neuronal nuclear antigen) staining revealed neurodegeneration in the CA3 region (arrows) of a *gl* hippocampus at P47. Scale bar, 100 μm .

an *agouti* background, the fur of *Clcn7*^{-/-} mice is grey (data not shown), just like the coat colour of grey-lethal mice^{5,23}. The disruption of CIC-7 leads not only to osteopetrosis³, but also to retinal³ and central nervous system⁴ degeneration, which are related to lysosomal storage disease⁴. We detected similar retinal and hippocampal degeneration in grey-lethal mice (Fig. 4h–j). Like *Clcn7*^{-/-} mice⁴, they showed electron-dense storage material in neurons and renal proximal tubular cells (Supplementary Fig. S5). Again similar to *Clcn7*^{-/-} mice⁴, the pH of gl lysosomes seemed normal (Supplementary Fig. S6), possibly pointing to altered acidification during lysosome formation, or to a role for lysosomal chloride^{4,18}. Taken together, we suggest that the CIC-7–Ostm1 complex is also important for the function of melanocytes and lysosomes. Patients with *OSTM1* mutations may develop lysosomal storage disease in addition to osteopetrosis.

Our work has identified Ostm1 as a hitherto unknown ancillary β -subunit of CIC-7. Ostm1 requires CIC-7 to travel to lysosomes, whereas CIC-7 can reach its destination without Ostm1. The stability of CIC-7 depends on its association with Ostm1. As pronounced glycosylation is thought to protect lysosomal membrane proteins from degradation^{24,25}, one might speculate that the highly glycosylated Ostm1 protein shields CIC-7, the only mammalian CLC protein lacking N-linked glycosylation sites, from lysosomal proteases. The osteopetrosis, lysosomal storage and neurodegeneration observed upon loss of Ostm1 may be entirely explained by a large reduction in CIC-7 protein levels.

METHODS

Please refer to Supplementary Methods for full details.

Mice. Grey-lethal mice^{5,26} obtained from Jackson Laboratory as double-heterozygous *GL/Le Edar*^{dl-J}/+ *Ostm1*^{gl/J} mice were crossed once to C57BL/6 mice and further bred to isolate the *gl* allele and produce homozygous *gl/gl* mice. Details of *Clcn7*^{-/-} mice have been published³.

Antibodies. Antibodies against the $\alpha 3$ subunit of the V-type H⁺-ATPase (peptide PDASTLENSWSPDEEK) or Ostm1 (LKSSTSFANIQENAT) were raised in guinea pigs and rabbits, affinity purified and tested for specificity on *oc* and *gl* knockout tissue, respectively^{5,11}. Antibodies against CIC-3 and CIC-7 have been described^{3,9}. See Supplementary Information for details of other antibodies.

DNA constructs. The *Ostm1* cDNA was amplified by polymerase chain reaction with reverse transcription (RT-PCR) from mouse kidney, and cloned into an eukaryotic expression vector. A mutant mimicking the OSTM1 fragment remaining in an osteopetrosis patient⁵ was generated by replacing the sequence encoding Ostm1 amino acids 266–338 with an HA-epitope (resulting sequence VEDA-VD-YPYDVPDYA-stop).

Cell culture. Fibroblasts from wild-type or *Clcn7*^{-/-} mice, HEK293 or HeLa cells were prepared and cultured as described⁴. For protease inhibition, fibroblasts were cultured for 24 h in the presence of 20 μ M leupeptin (Roche) or 10 μ M E64 (Sigma).

Immunohistochemistry and histology. Immunostaining of cryo- and paraffin sections and histology were done as described^{3,4}.

Biochemical methods. Cellular membranes from mouse brain were fractionated by sedimentation velocity centrifugation as detailed in the Supplementary Methods. Membrane preparations or gradient fractions were deglycosylated using PNGaseF or EndoH (Roche). For immunoprecipitation experiments, Ostm1 or CIC-7 antibodies were crosslinked to protein A sepharose by dimethylpimelidate. Brain membranes were pelleted and solubilized in lysis buffer containing 1% Triton X-100. Non-solubilized material was removed by a 70,000g spin. After incubation with protein A sepharose–antibody complexes for 2 h at 4°C and washing, samples were eluted at pH 2.8, neutralized, and denatured in SDS sample buffer.

Received 25 August; accepted 16 December 2005.

- Jentsch, T. J., Neagoe, I. & Scheel, O. CLC chloride channels and transporters. *Curr. Opin. Neurobiol.* **15**, 319–325 (2005).
- Jentsch, T. J., Poët, M., Fuhrmann, J. C. & Zdebik, A. A. Physiological functions

of CLC Cl⁻ channels gleaned from human genetic disease and mouse models. *Annu. Rev. Physiol.* **67**, 779–807 (2005).

- Kornak, U. et al. Loss of the CIC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* **104**, 205–215 (2001).
- Kasper, D. et al. Loss of the chloride channel CIC-7 leads to lysosomal storage disease and neurodegeneration. *EMBO J.* **24**, 1079–1091 (2005).
- Chalhoub, N. et al. Grey-lethal mutation induces severe malignant autosomal recessive osteopetrosis in mouse and human. *Nature Med.* **9**, 399–406 (2003).
- Rajapurohitam, V. et al. The mouse osteopetrotic grey-lethal mutation induces a defect in osteoclast maturation/function. *Bone* **28**, 513–523 (2001).
- Fischer, T., De Vries, L., Meerloo, T. & Farquhar, M. G. Promotion of Gai3 subunit down-regulation by GIPN, a putative E3 ubiquitin ligase that interacts with RGS-GAIP. *Proc. Natl Acad. Sci. USA* **100**, 8270–8275 (2003).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Stobrawa, S. M. et al. Disruption of CIC-3, a chloride channel expressed on synaptic vesicles, leads to a loss of the hippocampus. *Neuron* **29**, 185–196 (2001).
- Brandt, S. & Jentsch, T. J. CIC-6 and CIC-7 are two novel broadly expressed members of the CLC chloride channel family. *FEBS Lett.* **377**, 15–20 (1995).
- Scimeca, J. C. et al. The gene encoding the mouse homologue of the human osteoclast-specific 116-kDa V-ATPase subunit bears a deletion in osteosclerotic (*oc/oc*) mutants. *Bone* **26**, 207–213 (2000).
- Nishi, T. & Forgac, M. Molecular cloning and expression of three isoforms of the 100-kDa a subunit of the mouse vacuolar proton-translocating ATPase. *J. Biol. Chem.* **275**, 6824–6830 (2000).
- Li, Y. P., Chen, W., Liang, Y., Li, E. & Stashenko, P. *Atp6i*-deficient mice exhibit severe osteopetrosis due to loss of osteoclast-mediated extracellular acidification. *Nature Genet.* **23**, 447–451 (1999).
- Frattini, A. et al. Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis. *Nature Genet.* **25**, 343–346 (2000).
- Kornak, U. et al. Mutations in the $\alpha 3$ subunit of the vacuolar H⁺-ATPase cause infantile malignant osteopetrosis. *Hum. Mol. Genet.* **9**, 2059–2063 (2000).
- Accardi, A. & Miller, C. Secondary active transport mediated by a prokaryotic homologue of CIC Cl⁻ channels. *Nature* **427**, 803–807 (2004).
- Piccolo, A. & Pusch, M. Chloride/proton antiporter activity of mammalian CLC proteins CIC-4 and CIC-5. *Nature* **436**, 420–423 (2005).
- Scheel, O., Zdebik, A., Lourdel, S. & Jentsch, T. J. Voltage-dependent electrogenic chloride proton exchange by endosomal CLC proteins. *Nature* **436**, 424–427 (2005).
- Quarello, P. et al. Severe malignant osteopetrosis caused by a *GL* gene mutation. *J. Bone Miner. Res.* **19**, 1194–1199 (2004).
- Ramírez, A. et al. Identification of a novel mutation in the coding region of the grey-lethal gene *OSTM1* in human malignant infantile osteopetrosis. *Hum. Mutat.* **23**, 471–476 (2004).
- Cleiren, E. et al. Albers-Schönberg disease (autosomal dominant osteopetrosis, type II) results from mutations in the *CICN7* chloride channel gene. *Hum. Mol. Genet.* **10**, 2861–2867 (2001).
- Frattini, A. et al. Chloride channel *CICN7* mutations are responsible for severe recessive, dominant, and intermediate osteopetrosis. *J. Bone Miner. Res.* **18**, 1740–1747 (2003).
- Boyce, B. F. Bad bones, grey hair, one mutation. *Nature Med.* **9**, 395–396 (2003).
- Fukuda, M. Lysosomal membrane glycoproteins. Structure, biosynthesis, and intracellular trafficking. *J. Biol. Chem.* **266**, 21327–21330 (1991).
- Eskelinen, E. L., Tanaka, Y. & Saftig, P. At the acidic edge: emerging functions for lysosomal membrane proteins. *Trends Cell Biol.* **13**, 137–145 (2003).
- Vacher, J. & Bernard, H. Genetic localization and transmission of the mouse osteopetrotic grey-lethal mutation. *Mamm. Genome* **10**, 239–243 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Schweizer for electron microscopy and retina morphology, S. Bauer, N. Krönke and E. Orthey for technical assistance, and R. Pohlmann for cathepsin D antiserum. This work was supported by the Deutsche Forschungsgemeinschaft, and by a fellowship from the Boehringer Ingelheim Fonds to L. Wartosch.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J.J. (Jentsch@zmn.uni-hamburg.de).

LETTERS

Flagellar motility is required for the viability of the bloodstream trypanosome

Richard Broadhead^{1*}, Helen R. Dawe^{2*}, Helen Farr^{2*}, Samantha Griffiths^{2*}, Sarah R. Hart^{3*}, Neil Portman^{2*}, Michael K. Shaw², Michael L. Ginger², Simon J. Gaskell³, Paul G. McKean¹ & Keith Gull²

The 9+2 microtubule axoneme of flagella and cilia represents one of the most iconic structures built by eukaryotic cells and organisms. Both unity and diversity are present among cilia and flagella on the evolutionary as well as the developmental scale. Some cilia are motile, whereas others function as sensory organelles and can variously possess 9+2 and 9+0 axonemes and other associated structures¹. How such unity and diversity are reflected in molecular repertoires is unclear. The flagellated protozoan parasite *Trypanosoma brucei* is endemic in sub-Saharan Africa, causing devastating disease in humans and other animals². There is little hope of a vaccine for African sleeping sickness and a desperate need for modern drug therapies³. Here we present a detailed proteomic analysis of the trypanosome flagellum. RNA interference (RNAi)-based interrogation of this proteome provides functional insights into human ciliary diseases and establishes that flagellar function is essential to the bloodstream-form trypanosome. We show that RNAi-mediated ablation of various proteins identified in the trypanosome flagellar proteome leads to a rapid and marked failure of cytokinesis in bloodstream-form (but not procyclic insect-form) trypanosomes, suggesting that impairment of flagellar function may provide a method of disease control. A postgenomic meta-analysis, comparing the evolutionarily ancient trypanosome with other eukaryotes including humans, identifies numerous trypanosome-specific flagellar proteins, suggesting new avenues for selective intervention.

Flagellum-mediated migration between the gut and salivary glands of its tsetse fly vector is essential for progression of the trypanosome life cycle. However, the necessity for motility in an extracellular bloodstream-form trypanosome is unclear. In both forms, a single attached flagellum emerges from a posterior flagellar pocket (Fig. 1a) and comprises a membrane-bound axoneme and associated paraflagellar rod (PFR; Fig. 1b). We isolated the structural axoneme and associated PFR and basal body from procyclic trypanosomes by a well-characterized procedure of detergent and high-salt treatment⁴. Bands and spots were cut from one- (Fig. 1c) and two-dimensional gels and digested with trypsin, and the resulting peptides analysed by reverse-phase high-performance liquid chromatography coupled with tandem mass spectrometry. Bands and spots excised from ten representative gels were analysed, resulting in the identification of 522 nonredundant proteins.

Further manual mass spectrometric validation confirmed 380 proteins (see Supplementary Methods and Supplementary Fig. 1 for peptide numbers and coverage). Highly basic proteins are known to contaminate microtubule preparations owing to charge interactions⁴, and we noted some highly basic ribosomal proteins as recognizable contaminants. To limit contamination, we filtered the data by using an isoelectric point (pI) value of 10.2 as a cut-off (the pI

of the most acidic ribosomal protein), and placed 49 proteins (30 of which were ribosomal proteins) in a separate pool undoubtedly containing some genuine flagella components (for example, TbDIP13). The remaining 331 proteins constitute a *T. brucei* flagellum proteome (TbFP) characterized by both the inclusion of many known flagellar proteins and a lack of proteins from other subcellular compartments.

We subjected the TbFP to an *in silico* screen, testing which of these 331 proteins (and the pI-filtered set) were encoded in genomes of flagellated and non-flagellated eukaryotes (Supplementary Fig. 1). The TbFP is characterized by the absence of homologues in the genomes of non-flagellated eukaryotes (land plants, fungi and red algae; Fig. 1d). Shared components, such as α - or β -tubulin, are expected. We found that many TbFP proteins have homologues in the related trypanosomatid parasites *Trypanosoma cruzi* (312) and *Leishmania major* (286). Conservation of homologues between trypanosomatids and other flagellates reflects known structural differences among flagella of evolutionarily divergent organisms, being highest in *Tetrahymena*, human and *Chlamydomonas* (Fig. 1d). We found a much reduced proportion of homologues shared with *Plasmodium*, *Caenorhabditis* and *Drosophila*, which build motile or sensory cilia that are divergent in aspects of axoneme or basal body structure or formation⁵. We found that 208 TbFP proteins (Fig. 1d) are trypanosomatid-specific and probably represent organism-specific flagellar structures and functions. Although these most obviously include the PFR, some are likely to be axonemal.

Several studies have sought to identify components of the eukaryotic flagellum by proteomic^{6–9} or bioinformatic^{10,11} strategies. Although highly informative, bioinformatic and comparative genomic approaches exclude genuine conserved flagella components that have homologues and orthologues in non-flagellates. Similarly, occurrence of a protein in proteomes reflects the biochemical nature of the purified material. The TbFP, the *Tetrahymena thermophila* cilium⁹ and the axonemal fraction of the *Chlamydomonas* flagellar proteome⁸ are data sets derived from comparable preparations of insoluble flagellar architectures, with shared and organism-specific structures. To carry out a meta-analysis of these three data sets (Fig. 1e), we devised a stringent reciprocal BLASTP protocol that necessitated reanalysis of each data set to acquire the most current gene models available. After unification by applying the pI filter to all data sets, we undertook three reciprocal pairwise comparisons. We found that the TbFP contains the largest number of proteins in any of the three data sets. The presence of the PFR, an extra-axonemal, trypanosome-specific structure might explain the large number of proteins not encoded in the genomes of the other two organisms (indeed, the TbFP contains all of the known PFR proteins). However,

¹Department of Biological Sciences, Lancaster University, Lancaster LA1 4YQ, UK. ²Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK. ³Michael Barber Centre for Mass Spectrometry, School of Chemistry, University of Manchester, Sackville Street, Manchester M60 1QD, UK.

*These authors contributed equally to this work.

the *Chlamydomonas* and *Tetrahymena* pairwise axoneme-only comparisons revealed correspondingly large cohorts not represented in the genome of the opposing organism. Thus, the structural homogeneity of the 9+2 axoneme seems to be dependent on a relatively few evolutionarily conserved proteins and masks considerable organism-specific elaborations that go beyond the presence of obvious extra-axonemal structures.

Present in the conserved subset are proteins with implications for human inherited diseases. Ciliary diseases are characterized by clinical presentations such as retinal degeneration, primary ciliary dyskinesia, hydrocephalus, polycystic kidney disease and polydactyly¹. In addition to all three axonemal dyneins known to cause primary ciliary dyskinesia¹², interrogation of the TbFP identified homologues of a further seven proteins directly implicated in diseases of human, mouse or zebrafish (Supplementary Fig. 2a). This list includes Hydin¹³, PACRG¹⁴ and Scorpion¹⁵; two of these, TbHydin and TbPACRG, are validated functionally here for the first time to our knowledge. We determined the human chromosomal locus for every homologue represented in the TbFP, which identified 34 genes mapping to 25 loci where diseases with a clinical spectrum suggestive of ciliary dysfunction have been mapped genetically, but where the causal gene has not been identified. We suggest that these 34 genes (Supplementary Fig. 2b) represent candidates underlying syndromes including primary ciliary dyskinesia, polycystic kidney

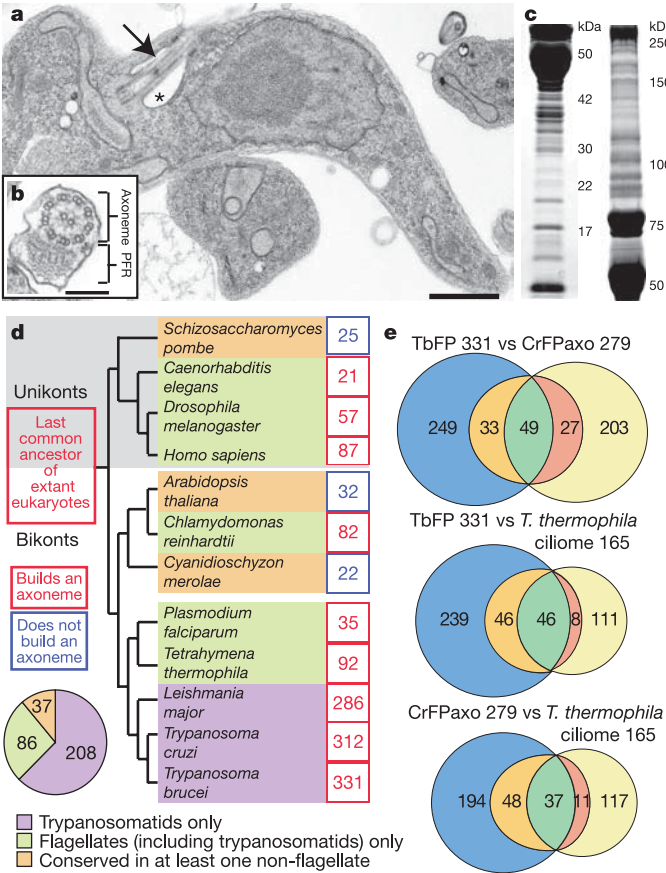


Figure 1 | The *T. brucei* flagellar proteome. **a**, Transmission electron micrograph showing the flagellum (arrow) and flagellar pocket (asterisk). **b**, The 9+2 axoneme and the PFR. **c**, One-dimensional electrophoretic separation of isolated flagella. **d**, Distribution of TbFP protein homologues. **e**, Comparison of the *T. brucei*, *Chlamydomonas reinhardtii* (CrFPaxo) and *T. thermophila* flagellar proteomes. For each comparison, green indicates present in both; orange/red indicates present in opposing genome but not opposing proteome; blue/yellow indicates no homologue in opposing genome. Homology determined by alignment to each genome with reciprocal alignment by BLASTP. Scale bars, 1 μm (**a**); 200 nm (**b**).

disease, macular and cone-rod dystrophies, retinitis pigmentosa, Rieger and BRESEK (see <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM> for syndrome details).

We subjected the orthologue of the cryptic hydrocephalus-inducing protein Hydin to functional analysis by using inducible RNAi in procyclic trypanosomes. This resulted in a severe motility defect, revealing an impairment of ciliary function as a likely explanation for the aetiology of this disease model (Fig. 2). A further ten TbFP proteins were interrogated by this RNAi approach (Fig. 2a), eight of which (including the trypanosome-specific proteins PFR2 and HERTS) showed a flagellar phenotype on ablation (Fig. 2a). All mutants were viable except one in which flagellar detachment caused pleiotropic effects¹⁶. Two genes (*TbPACRGA* and *TbPACRGB*) comprise a gene family in *T. brucei* and gave no phenotype on individual ablation. Simultaneous knockdown, however, produced paralysed flagella, suggesting functional redundancy.

The high success rate of the RNAi interrogation provides experimental evidence for the integrity of the TbFP. These phenotypes extend the catalogue of trypanosome flagellar RNAi mutants, which have all been made in the procyclic form^{16–19}, and show that motility can be severely compromised while proliferation remains unaffected. We extended our functional interrogation by carrying out RNAi against five proteins in our TbFP validation set in bloodstream-form trypanosomes (Supplementary Fig. 3a). Unexpectedly, we found that all five RNAi analyses resulted in bloodstream-form trypanosomes that did not complete cytokinesis, yielding monstrous cells with an inability to proliferate (Fig. 3 and Supplementary Fig. 3b–h). This striking contrast to the viable RNAi phenotypes produced in

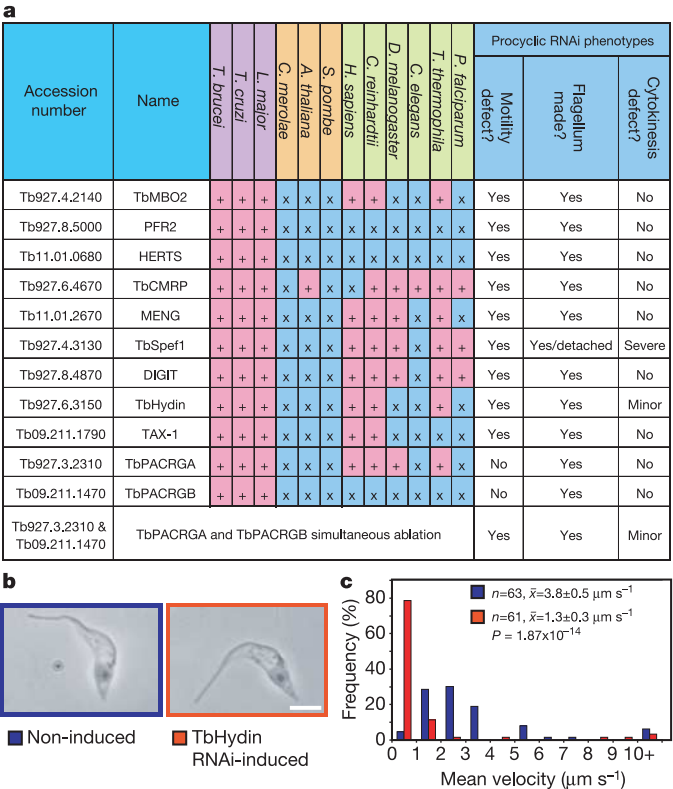


Figure 2 | Biological verification of the *T. brucei* flagellar proteome. **a**, Biological validation of the TbFP by RNAi of procyclic trypanosomes, showing phylogenetic distribution and RNAi phenotype. ‘+’ indicates presence of homologues; ‘x’ indicates absence of homologues. **b**, **c**, Phenotype of *T. brucei* Hydin (TbHydin). On RNAi induction the flagellum is built (**b**), but motility is severely compromised in individual cells and the cell population (**c**). Scale bar, 5 μm (**b**). Blue indicates non-induced control; red indicates TbHydin RNAi-induced. *P* value determined by Mann–Whitney *U*-test. Population statistic given as the mean $\pm$ s.e.m.

procyclic cells is shown in Fig. 3a–c. In each case, RNAi induction led to an inhibition of cytokinesis.

Trypanosomes have one kinetoplast (mitochondrial DNA) and one nucleus (1K1N), and replicate these once during a cell cycle²⁰; on RNAi induction, however, bloodstream cells ceased dividing but continued to progress through the cell cycle. Cells initiated new rounds of S phase and mitosis, leading to large contorted cells containing multiple kinetoplasts and nuclei (Fig. 3d, i, j). At 24 h after induction, the RNAi-induced population of ‘monsters’ contained large numbers of 4K4N and 8K8N cells (Fig. 3d).

Included in the bloodstream RNAi set were newly identified flagellar proteins such as TAX-1, TbPACRGA and DIGIT in addition to an orthologue (TbMBO2) of a *Chlamydomonas* protein known to regulate motility²¹. Given the implications of these observations, we tested whether the well-characterized procyclic *snl* mutant phenotype^{17,18} (paralysed but viable) of the *T. brucei* specific PFR2 protein was different in the bloodstream form. We found that this mutant also produced a rapid, lethal and unusual phenotype with production of monstrous cells. Notably, the same construct expressed in procyclic cells reproduced the published paralysed, but viable, phenotype (Fig. 2a).

A very particular phenotype led to the death of bloodstream cells. Uninduced cells had a normal morphology (Fig. 3e), producing a new flagellum during the cell cycle before cytokinesis, with the flagellum emerging from the flagellar pocket and extending along the trypanosome. On RNAi induction, cells produced new flagella but did not complete division and lost all normal morphogenetic axes, becoming monstrously contorted and multiflagellated (Fig. 3f–h). Thin-section electron microscopy showed that the cell periphery was highly convoluted; multiple nuclei were present and some flagella were attached to the outside of the cell, while others were

located in a highly enlarged flagellar pocket (Fig. 3i, j and Supplementary Fig. 3b, c, f, g). There were also many flagella with two axonemes, which could be either parallel or antiparallel, indicative of a total loss of morphogenetic patterning (Supplementary Fig. 3d, e). These characteristics are also found in the TbMBO2 and DIGIT phenotypes (Supplementary Fig. 3f, g). Examination of the flagellar pockets indicated that although large they still showed clathrin-coated pits, suggesting that endocytotic processes were operating to some extent (for example, in the TAX-1 phenotype; Supplementary Fig. 3h).

The explanation for the lethal phenotype in bloodstream cells is that cytokinesis fails as a primary event in the absence of correct flagellar motility. Subsequent rounds of the cell cycle compound these events as new flagella and pockets are formed at inappropriate locations. The lack of precise morphogenetic axes in the resulting ‘monstrous cells’ leads to pleiotropic effects compromising membrane–cytoskeletal balance during morphogenesis of the pocket structure. Examination of endocytosis shows that the multiple pockets are active and can facilitate entry of antibodies and lectins. However, there is much variation in their individual capacity for vectorial internalization to nearby endomembrane compartments (Supplementary Fig. 4). An imbalance in the known high flux of endocytosis in the bloodstream form²², coupled with a focus of the secretory pathway on particular pockets in a multiflagellated cell, will lead to the large flagellar pocket phenotype as a secondary event.

In summary, we have shown that flagellum function cannot be compromised and is essential in the bloodstream trypanosome. The severity of the phenotype, its rapid onset, its lethality and its occurrence after the ablation of proteins of diverse function and location in either the axoneme or the trypanosome-specific PFR is highly significant. Coupled with our identification of a set of trypanosome-specific proteins in the TbFP, this suggests that impair-

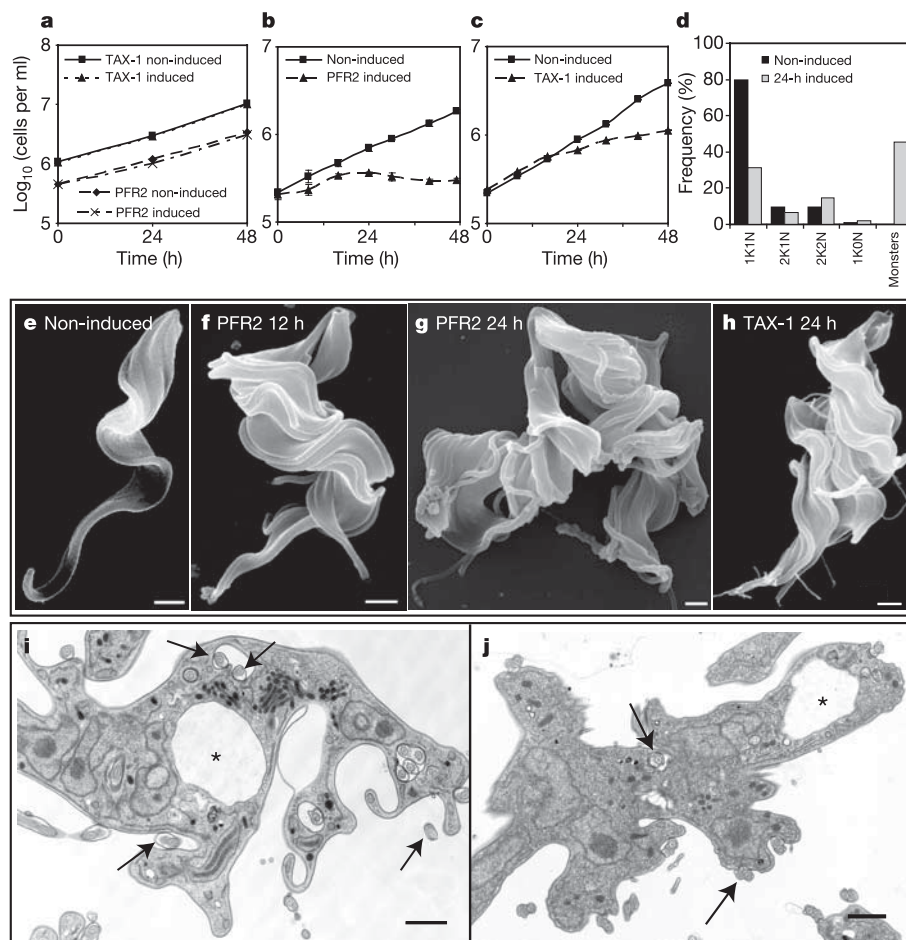


Figure 3 | Flagellar motility is essential for bloodstream trypanosomes. **a–c**, Ablation of PFR2 and TAX-1 affects growth in bloodstream (**b**, **c**), but not procyclic (**a**), forms. **d**, Numbers of cells with multiple nuclei and kinetoplasts increase after ablation of TAX-1 in bloodstream forms. **e–h**, Scanning electron micrograph showing bloodstream-form morphogenesis before (**e**) and after (**f–h**) ablation of PFR2 and TAX-1 flagellar proteins. **i**, **j**, Transmission electron microscopy analysis of PFR2 (**i**) and TAX-1 (**j**) ablation in bloodstream forms. Arrows denote axonemes; asterisks denote flagellar pockets (**i**, **j**). Scale bar, 1 μ m.

ment of flagellar function may provide an avenue for disease control in African sleeping sickness.

METHODS

Preparation of flagella proteins for mass spectrometry. Procytic *T. brucei* cells were collected by centrifugation (800g, 10 min), washed once in PBS containing 5 μ M E-64d, and resuspended in PEME (100 mM PIPES, 2 mM EGTA, 0.1 mM EDTA and 1 mM MgSO₄; pH 6.9) containing 1% Nonidet P40, 1× Focus protease-arrest mix (Calbiochem), 7.5 μ M Pepstatin A and 5 μ M E-64d. Cytoskeletons were collected by centrifugation and subjected to two rounds of salt extraction with PEME containing 1 M NaCl, 200 μ g ml⁻¹ of DNaseI, 50 μ g ml⁻¹ RNaseA, 1× Focus protease-arrest mix, 7.5 μ M Pepstatin A and 5 μ M E-64d at 4 °C for 10 min. Flagella pellets were collected by centrifugation (16,000g, 15 min, 4 °C), washed twice in PEME and resuspended in Laemmli buffer²³. Proteins from 6 × 10⁸ cell equivalents were resolved by one-dimensional SDS-PAGE (using standard protocols) and two-dimensional SDS-PAGE (see Supplementary Methods) and visualized by staining with Coomassie blue. Regions, bands and spots were excised, digested with trypsin and analysed by mass spectrometry. Verification of the goodness of fit of putative matches to predicted protein sequences was done manually (see Supplementary Methods). All data that yielded information used to construct the TbFP will be uploaded to the EBI Proteomics Identification database (<http://www.ebi.ac.uk/pride/>), a public data repository. This resource is freely available for use by the proteomics community.

Analysis of the TbFP. A detailed description of the reciprocal BLASTP²⁴ protocol used in comparative analysis is given in Supplementary Methods. Genetic loci containing human homologues of TbFP proteins were identified and compared to the following: first, loci implicated in known cilia-related diseases and diseases of unknown cause where the clinical presentation suggests a defect in cilia or flagella function (derived from OMIM: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM&tool=toolbar>); second, the RetNet database of retinal disorders (<http://www.sph.uth.tmc.edu/Retnet/>); third, human orthologues of mouse infertility models²⁵; and last, genes that cause polycystic kidney disease in zebrafish¹⁵.

Trypanosome cell culture and RNAi. Procytic (427, 29-13 and derivatives) and bloodstream (90-13 and derivatives) cell lines were maintained and transfected by standard methods²⁶. Cell growth was monitored by a CASY1 cell-counter and analyser system (Schärfe System GmbH). RNAi studies were carried out with 400–600 base-pair fragments of *T. brucei* coding sequence, which were amplified by PCR from genomic DNA and cloned into the inducible RNAi vector p2T7-177 (ref. 27). Primer sequences are given in Supplementary Methods. For counts of nuclei and kinetoplast number, cells were fixed in 3.6% paraformaldehyde in PBS and dried onto microscope slides. 4,6-Diamidino-2-phenylindole (DAPI) was used to visualize DNA. We counted 500 cells at each time point. Motility analyses were done essentially as described²⁸: time-lapse sequences were captured every 2 s for 40 s by using phase-contrast microscopy at ×10 magnification and 22 °C. Mean velocities were determined for individual cells and for each cell line collectively. Statistical significance was tested by a two-tailed Mann–Whitney *U*-test because samples varied in size and distributions were not suited to parametric testing. All analyses showed significance at $\alpha = 0.01$ with the exception of TbPACRGA and TbPACRGB single knockdowns, which were not significant at this level.

Electron microscopy. For thin-section electron microscopy, cells were fixed in 2.5% glutaraldehyde, 2% paraformaldehyde and 0.1% picric acid in 100 mM phosphate (pH 6.5) for 2 h at 4 °C followed by post-fixation in 1% osmium tetroxide in 100 mM phosphate buffer (pH 6.5) with 50 mM sucrose for 1 h at 4 °C. The fixed material was stained *en bloc* with 2% aqueous uranyl acetate for 2 h at 4 °C, dehydrated, and embedded in epon resin. For scanning electron microscopy, samples were prepared essentially as described²⁹.

Received 8 November; accepted 14 December 2005.

1. Afzelius, B. A. Cilia-related diseases. *J. Pathol.* **204**, 470–477 (2004).
2. Sternberg, J. M. Human African trypanosomiasis: clinical presentation and immune response. *Parasite Immunol.* **26**, 469–476 (2004).
3. Steverding, D. & Tyler, K. M. Novel antitrypanosomal agents. *Expert Opin. Invest. Drugs* **14**, 939–955 (2005).
4. Robinson, D., Beattie, P., Sherwin, T. & Gull, K. Microtubules, tubulin, and microtubule-associated proteins of trypanosomes. *Methods Enzymol.* **196**, 285–299 (1991).
5. Briggs, L. J., Davidge, J. A., Wickstead, B., Ginger, M. L. & Gull, K. More than one way to build a flagellum: comparative genomics of parasitic protozoa. *Curr. Biol.* **14**, R611–R612 (2004).
6. Ostrowski, L. E. *et al.* A proteomic analysis of human cilia: identification of novel components. *Mol. Cell. Proteomics* **1**, 451–465 (2002).

7. Padma, P. *et al.* Identification of a novel leucine-rich repeat protein as a component of flagellar radial spoke in the Ascidian *Ciona intestinalis*. *Mol. Biol. Cell* **14**, 774–785 (2003).
8. Pazour, G. J., Agrin, N., Leszyk, J. & Witman, G. B. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170**, 103–113 (2005).
9. Smith, J. C., Northey, J. G., Garg, J., Pearlman, R. E. & Siu, K. W. Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of *Tetrahymena thermophila*. *J. Proteome Res.* **4**, 909–919 (2005).
10. Avidor-Reiss, T. *et al.* Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* **117**, 527–539 (2004).
11. Li, J. B. *et al.* Comparative genomics identifies a flagellar and basal body proteome that includes the BBS human disease gene. *Cell* **117**, 541–552 (2004).
12. Chodhari, R., Mitchison, H. M. & Meeks, M. Cilia, primary ciliary dyskinesia and molecular genetics. *Paediatr. Respir. Rev.* **5**, 69–76 (2004).
13. Davy, B. E. & Robinson, M. L. Congenital hydrocephalus in *hy3* mice is caused by a frameshift mutation in *Hydin*, a large novel gene. *Hum. Mol. Genet.* **12**, 1163–1170 (2003).
14. Lorenzetti, D., Bishop, C. E. & Justice, M. J. Deletion of the Parkin coregulated gene causes male sterility in the *quaking(viable)* mouse mutant. *Proc. Natl Acad. Sci. USA* **101**, 8402–8407 (2004).
15. Sun, Z. *et al.* A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney. *Development* **131**, 4085–4093 (2004).
16. Moreira-Leite, F. F., Sherwin, T., Kohl, L. & Gull, K. A trypanosome structure involved in transmitting cytoplasmic information during cell division. *Science* **294**, 610–612 (2001).
17. Bastin, P., Ellis, K., Kohl, L. & Gull, K. Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. *J. Cell Sci.* **113**, 3321–3328 (2000).
18. Bastin, P., Sherwin, T. & Gull, K. Paraflagellar rod is vital for trypanosome motility. *Nature* **391**, 548 (1998).
19. Hill, K. L. Biology and mechanism of trypanosome cell motility. *Eukaryot. Cell* **2**, 200–208 (2003).
20. Woodward, R. & Gull, K. Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J. Cell Sci.* **95**, 49–57 (1990).
21. Tam, L. W. & Lefebvre, P. A. The *Chlamydomonas* MBO2 locus encodes a conserved coiled-coil protein important for flagellar waveform conversion. *Cell Motil. Cytoskeleton* **51**, 197–212 (2002).
22. Field, M. C. & Carrington, M. Intracellular membrane transport systems in *Trypanosoma brucei*. *Traffic* **5**, 905–913 (2004).
23. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
24. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
25. Liska, F. Selected genetic aspects of male infertility—what animal models tell us. *Folia Biol. (Praha)* **49**, 129–141 (2003).
26. Wirtz, E. & Clayton, C. Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science* **268**, 1179–1183 (1995).
27. Wickstead, B., Ersfeld, K. & Gull, K. Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **125**, 211–216 (2002).
28. Gadelha, C., Wickstead, B., de Souza, W., Gull, K. & Cunha-e-Silva, N. Cryptic paraflagellar rod in endosymbiont-containing kinetoplastid protozoa. *Eukaryot. Cell* **4**, 516–525 (2005).
29. Sherwin, T. & Gull, K. The cell division cycle of *Trypanosoma brucei*: timing of event markers and cytoskeletal modulations. *Phil. Trans. R. Soc. Lond. B* **323**, 573–588 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the genome centres, particularly GeneDB at the Sanger Institute, and our laboratory colleagues, especially S. Gordon and P. Taylor, for access to unpublished data and constructive discussions. This work was funded by the Biotechnology and Biological Sciences Research Council, the Wellcome Trust, the Royal Society and the Edward P. Abraham Trust. M.L.G. is a Royal Society University Research Fellow; K.G. is a Wellcome Trust Principal Research Fellow.

Author Contributions The first six authors contributed equally to this work and are listed alphabetically. Individual contributions were as follows: R.B. proteomics and bioinformatic analysis; H.R.D. proteomics, bioinformatic and functional analyses; H.F. and S.G. functional analysis; S.R.H. proteomics; N.P. proteomics, bioinformatic and functional analyses.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.G. (keith.gull@path.ox.ac.uk).

LETTERS

Cryopyrin activates the inflammasome in response to toxins and ATP

Sanjeev Mariathasan¹, David S. Weiss⁴, Kim Newton¹, Jacqueline McBride³, Karen O'Rourke¹, Meron Roose-Girma², Wyne P. Lee³, Yvette Weinrauch⁵, Denise M. Monack⁴ & Vishva M. Dixit¹

A crucial part of the innate immune response is the assembly of the inflammasome, a cytosolic complex of proteins that activates caspase-1 to process the proinflammatory cytokines interleukin (IL)-1 β and IL-18. The adaptor protein ASC is essential for inflammasome function^{1,2}, binding directly to caspase-1 (refs 3, 4), but the triggers of this interaction are less clear. ASC also interacts with the adaptor cryopyrin (also known as NALP3 or CIAS1)^{5,6}. Activating mutations in cryopyrin are associated with familial cold autoinflammatory syndrome, Muckle-Wells syndrome and neonatal onset multisystem inflammatory disease, diseases that are characterized by excessive production of IL-1 β ^{5,7}. Here we show that cryopyrin-deficient macrophages cannot activate caspase-1 in response to Toll-like receptor agonists plus ATP, the latter activating the P2X₇ receptor to decrease intracellular K⁺ levels^{8,9}. The release of IL-1 β in response to nigericin, a potassium ionophore, and maitotoxin, a potent marine toxin, was also found to be dependent on cryopyrin. In contrast to *Asc*^{-/-} macrophages, cells deficient in the gene encoding cryopyrin (*Cias1*^{-/-}) activated caspase-1 and secreted normal levels of IL-1 β and IL-18 when infected with Gram-negative *Salmonella typhimurium* or *Francisella tularensis*. Macrophages exposed to Gram-positive *Staphylococcus aureus* or *Listeria monocytogenes*, however, required both ASC and cryopyrin to activate caspase-1 and secrete IL-1 β . Therefore, cryopyrin is essential for inflammasome activation in response to signalling pathways triggered specifically by ATP, nigericin, maitotoxin, *S. aureus* or *L. monocytogenes*.

Cryopyrin-deficient mice (Supplementary Fig. S1) were generated by gene targeting to investigate the role of cryopyrin in inflammatory responses to pathogen-derived molecules. Cryopyrin-deficient (*Cias1*^{-/-}) macrophages stimulated with the Toll-like receptor-4 (TLR4) agonist lipopolysaccharide (LPS) phosphorylated I κ B α and ERK normally (Fig. 1a), and they secreted normal amounts of TNF- α (Fig. 1b), IL-12 p40 (Fig. 1c), IL-6 and IL-10 (data not shown). Similar results were obtained using the TLR2 agonists Pam₃CSK₄ and heat-killed *L. monocytogenes* (HKLM) (data not shown). Our results show that cryopyrin is dispensable for NF- κ B signalling by TLR2 and TLR4 in macrophages.

Because mutant variants of cryopyrin are associated with diseases in which IL-1 β is produced in excess⁵⁻⁷, we measured IL-1 β released from *Cias1*^{-/-} macrophages treated with TLR agonists and ATP (Fig. 1d). TLR agonists induce pro-IL-1 β synthesis and ATP stimulates caspase-1-dependent cleavage and secretion of IL-1 β ¹⁰. In contrast to wild-type macrophages, which secreted readily detectable amounts of IL-1 β and IL-18 in response to ATP plus ultra-pure LPS, Pam₃CSK₄, HKLM, R848 (TLR7/8 agonist), or CpG oligonucleotides (TLR9 agonist), *Cias1*^{-/-} macrophages secreted negligible amounts

of these cytokines (Fig. 1d, e). As shown previously^{1,11}, *Asc*^{-/-} macrophages exhibited a similar defect in IL-1 β and IL-18 production (Fig. 1d, e). Macrophages from heterozygous *Cias1*^{+/-} mice secreted intermediate amounts of IL-1 β and IL-18. C3H/HeJ macrophages expressing a non-functional form of TLR4 (ref. 12) secreted IL-1 β and IL-18 in response to ATP plus either Pam₃CSK₄ or HKLM, but not LPS, demonstrating that our LPS was pure and not contaminated with other TLR agonists (Fig. 1d).

To determine whether IL-1 β secretion from *Cias1*^{-/-} macrophages was defective due to impaired pro-IL-1 β synthesis and/or impaired caspase-1 activation, we immunoprecipitated [³⁵S]-methionine-labelled pro-IL-1 β from LPS-primed macrophages. Wild-type, *Cias1*^{+/-} and *Cias1*^{-/-} macrophages produced comparable amounts of pro-IL-1 β (Fig. 1f, left panel), indicating that defective IL-1 β secretion from *Cias1*^{-/-} cells was not due to impaired pro-IL-1 β synthesis. Unlike their wild-type counterparts, however, *Cias1*^{-/-} macrophages did not cleave pro-IL-1 β after ATP treatment (Fig. 1f, right panel). This finding suggested that cryopyrin is essential for ATP-induced caspase-1 activation. A further indication of caspase-1 activation is its autocatalytic processing into p20 and p10 subunits. Western blotting for caspase-1 after LPS plus ATP treatment revealed the p10 and p20 subunits in wild-type but not *Cias1*^{-/-} macrophages (Fig. 1g). Thus, cryopyrin is essential for activation of caspase-1 in response to LPS plus ATP. Notably, ATP was necessary but not sufficient for caspase-1 activation (Fig. 1g). TLR signalling is probably needed for expression of essential inflammasome components. For example, LPS stimulation of TLR4 increases expression of caspase-11, and analyses of caspase-11-deficient mice and cells demonstrate that caspase-11 is essential for inflammasome function¹³.

To test whether the role of ATP in cryopyrin- and ASC-dependent caspase-1 activation relates to its ability to stimulate the P2X₇ receptor⁹ and thereby reduce intracellular K⁺ (ref. 8), we treated TLR-primed macrophages from wild-type, *Asc*^{-/-} and *Cias1*^{-/-} mice with nigericin or maitotoxin to deplete cytosolic K⁺ (refs 14, 15). Wild-type macrophages primed with LPS or Pam₃CSK₄ secreted IL-1 β and IL-18 in response to nigericin (Fig. 2a, b) or maitotoxin (Fig. 2c, d). By contrast, neither *Asc*^{-/-} nor *Cias1*^{-/-} macrophages released significant IL-1 β or IL-18. Our data therefore show that ASC and cryopyrin are essential for IL-1 β and IL-18 production by TLR-primed macrophages treated with agents that deplete intracellular K⁺. Neither nigericin nor maitotoxin alone induced IL-1 β release (Supplementary Fig. S2). Again, TLR signalling is probably required not only for the induction of pro-IL-1 β (Fig. 1f) but also for expression of other proteins that are essential for inflammasome function.

To determine whether cryopyrin is essential for inflammation

¹Molecular Oncology Department, ²Physiology Department, ³Immunology Department, Genentech Inc, 1 DNA Way, South San Francisco, California 94080, USA. ⁴Departments of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA. ⁵Department of Microbiology, New York University School of Medicine, New York, New York 10016, USA.

in vivo, wild-type and *Cias1*^{-/-} mice were injected with a lethal dose of LPS to induce caspase-1-dependent endotoxic shock¹⁶. All of the wild-type mice died within 48 h, whereas only ~30% of the *Cias1*^{-/-} mice had died after 72 h (Fig. 2e). Correlating with their enhanced survival, *Cias1*^{-/-} mice had markedly less serum IL-1 β and IL-18 than the wild-type mice (Fig. 2f). It is unclear why exogenous ATP is not required for IL-1 β and IL-18 secretion *in vivo*. We speculate that other cell types impacted by LPS *in vivo* provide the ATP needed to engage the P2X₇ receptor. Our results demonstrate that cryopyrin, like ASC, is also an important mediator of LPS-induced endotoxic shock¹.

A previous study suggested that bacterial muramyl dipeptide (MDP) can activate a cryopyrin-containing inflammasome¹⁷. MDP is also the ligand for NOD2, which resembles cryopyrin in that it possesses a nucleotide binding domain and leucine-rich repeats^{7,18–20}. Mutated NOD2 has been linked to Crohn's disease⁷. MDP induced phosphorylation of I κ B α and ERK in wild-type and *Cias1*^{-/-}

macrophages but, in agreement with published results¹⁹, *Nod2*^{-/-} macrophages were unresponsive (Fig. 3a). NOD2-deficiency, however, did not impact on LPS-induced phosphorylation of I κ B α and ERK, or subsequent I κ B α degradation (Fig. 3a). In terms of cytokine production, MDP enhanced LPS-induced secretion of TNF and IL-12 p40 by wild-type and *Cias1*^{-/-} macrophages but not *Nod2*^{-/-} macrophages (Fig. 3b, c). Because cryopyrin was essential for LPS- plus ATP-induced IL-1 β secretion, we could not assess whether MDP increased IL-1 β production in wild-type cells by engaging cryopyrin (Fig. 3d). Our data indicate that cryopyrin is dispensable for MDP-induced activation of NF- κ B and ERK, and confirm a crucial role for cryopyrin in ATP-induced activation of the macrophage inflammasome.

Next we determined whether cryopyrin is required for caspase-1 activation and IL-1 β release when macrophages are infected by specific bacterial pathogens. Unlike ASC, cryopyrin was dispensable for normal caspase-1 activation, IL-1 β secretion and macrophage

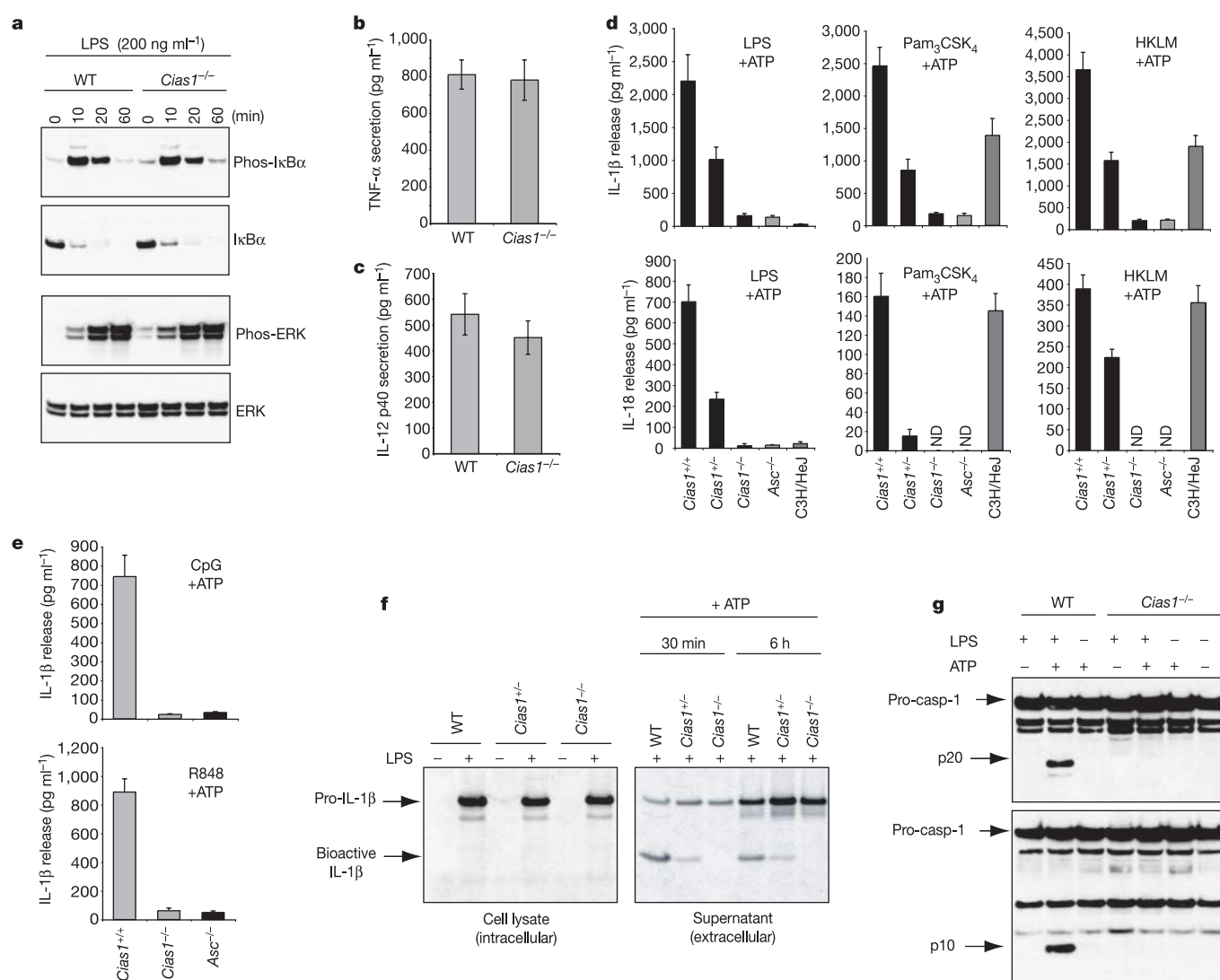


Figure 1 Cryopyrin is essential for caspase-1 activation and IL-1 β secretion in response to TLR agonists and ATP. **a**, Western blot analysis of phosphorylated and total I κ B α and ERK in wild-type (WT) and *Cias1*^{-/-} bone-marrow-derived macrophages stimulated with LPS. **b**, **c**, TNF (b) and IL-12 p40 (c) secretion by peritoneal macrophages cultured for 16 h with LPS. **d**, Macrophages primed for 16 h with LPS, Pam₃CSK₄, or HKLM were then pulsed with ATP. IL-1 β or IL-18 released in the next 3 h is shown. ND, not detected. **e**, IL-1 β secretion by macrophages primed with CpG DNA or R848 and then pulsed with ATP. Bars in **b–e** represent the mean $\pm$ s.d. of

triplicate wells. Results are representative of four independent experiments. **f**, Immunoprecipitation of [³⁵S]-labelled pro-IL-1 β and IL-1 β from macrophages treated with LPS (left panel). Pro-IL-1 β and IL-1 β secreted after subsequent ATP treatment were immunoprecipitated from culture supernatants (right panel). **g**, Western blot analysis of caspase-1 in peritoneal macrophages stimulated with LPS for 16 h and then pulsed with ATP for 20 min. All *in vitro* experiments were performed with ultra-pure LPS.

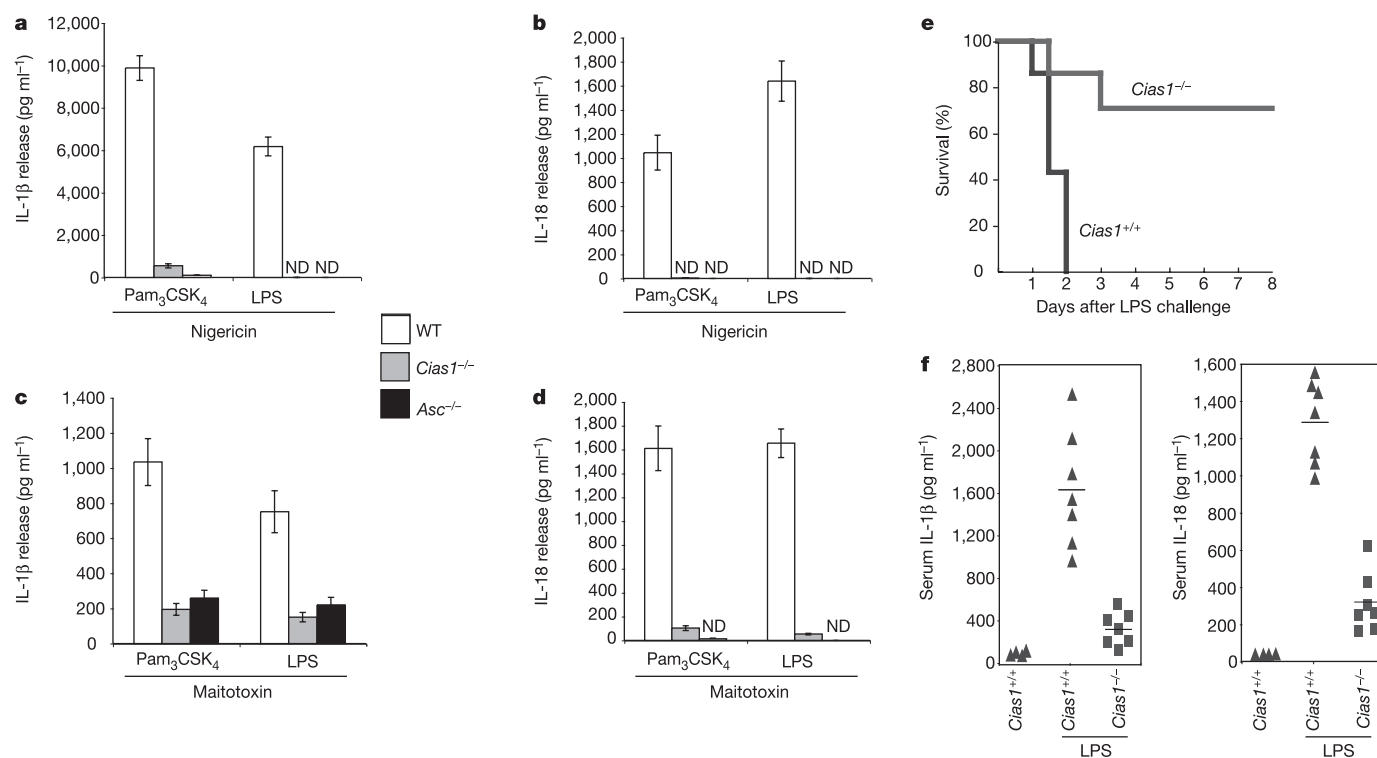


Figure 2 | Cryopyrin is essential for caspase-1 activation and IL-1 β secretion in response to TLR agonists plus nigericin or maitotoxin. **a–d**, Wild-type, *Cias1* $^{-/-}$ and *Asc* $^{-/-}$ macrophages were primed with Pam $_3$ CSK $_4$ or ultra-pure LPS and then treated with nigericin (**a**, **b**) or maitotoxin (**c**, **d**). IL-1 β (**a**, **c**) and IL-18 (**b**, **d**) release was measured by ELISA. Bars in **a–d** represent the mean $\pm$ s.d. of triplicate wells. Results are representative of three independent experiments. ND, not detected. **e**, Survival of 8-week-old

wild-type ($n = 7$) or *Cias1* $^{-/-}$ ($n = 7$) female mice injected intraperitoneally with crude LPS (40 mg kg $^{-1}$ of body weight) on day 0. **f**, Levels of IL-1 β and IL-18 in the serum of the mice in **e** at 3 h after injection. Lines indicate the mean serum level (IL-1 β : WT, 1,581 pg ml $^{-1}$; *Cias1* $^{-/-}$, 304 pg ml $^{-1}$, $P = 0.001$; IL-18: WT, 1,321 pg ml $^{-1}$; *Cias1* $^{-/-}$, 329 pg ml $^{-1}$, $P = 0.0004$).

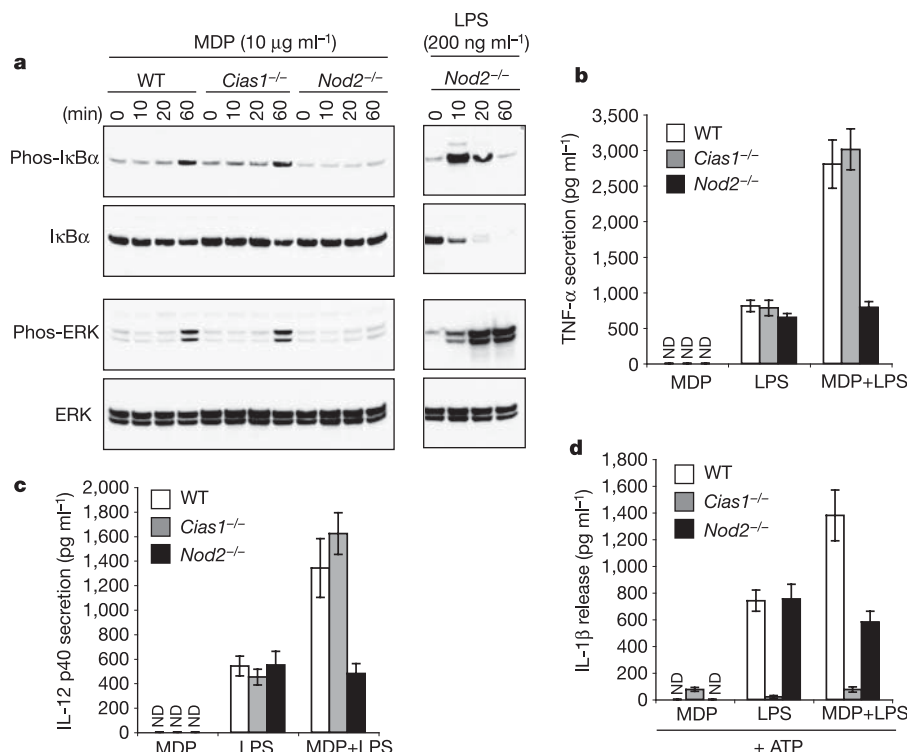


Figure 3 | MDP triggers NOD2-dependent, cryopyrin-independent NF- κ B and ERK signalling. **a**, Western blot analysis of phosphorylated or total I κ B α and ERK in wild-type, *Cias1* $^{-/-}$ and *Nod2* $^{-/-}$ macrophages stimulated with MDP or LPS. **b–d**, Secretion of TNF (**b**), IL-12 p40 (**c**), or

IL-1 β (**d**) by bone-marrow-derived macrophages treated with MDP and/or LPS. Macrophages in **d** were also pulsed with ATP. ND, not detected. Bars represent the mean $\pm$ s.d. of triplicate wells. Results are representative of three independent experiments.

cell death in response to Gram-negative *S. typhimurium* (Fig. 4a, b) or *F. tularensis* (Fig. 4c, d) (Supplementary Fig. S4). As shown previously^{1,21}, macrophages infected by *S. typhimurium* also require the adaptor protein Ipaf to activate caspase-1 (Fig. 4a). ASC and cryopyrin were essential for caspase-1 activation and secretion of IL-1 β or IL-18 when macrophages were cultured with Gram-positive *S. aureus* (Fig. 4e–g) or *L. monocytogenes* (Fig. 4h–l), but NOD2 was dispensable (data not shown). ASC or cryopyrin deficiency caused a specific defect in IL-1 β and IL-18 release because *Asc*^{-/-} and *Cias1*^{-/-} macrophages infected by *L. monocytogenes* and *S. aureus* yielded similar amounts of TNF to wild-type macrophages (Fig. 4k and data not shown). Macrophages cultured with *L. monocytogenes* deficient for the toxin listeriolysin O did not secrete IL-1 β (Fig. 4l), so we speculate that listeriolysin O perturbs intracellular K⁺ levels similar to extracellular ATP, nigericin and maitotoxin. Furthermore, live bacteria seem to be required because heat-killed *Listeria monocytogenes* (HKLM) induced very little IL-1 β secretion (data not shown). *S. aureus* deficient in alpha-, beta- or gamma-toxin induced comparable IL-1 β secretion to wild-type *S. aureus* (Supplementary

Fig. S5), indicating that other *S. aureus* toxins may contribute to caspase-1 activation.

Our results suggest that the caspase-1 inflammasome is a dynamic entity that is assembled from different adaptor proteins in a stimulus-dependent manner (Fig. 4m). We show that cryopyrin is essential for inflammasome activation in response to signalling pathways triggered by specific bacterial infections and to treatments that deplete intracellular K⁺. Ipaf and perhaps other members of the large NALP family of proteins might substitute for cryopyrin upon infection with Gram-negative bacteria such as *S. typhimurium* and *F. tularensis*. The cryopyrin-dependent response to extracellular ATP may represent a physiological response that occurs when ATP is released by dying cells and degranulating platelets. Future studies will need to address how activating mutations in cryopyrin circumvent the normal regulation of the inflammasome to produce inflammatory diseases such as familial cold autoinflammatory syndrome, Muckle-Wells syndrome and neonatal onset multisystem inflammatory disease^{22–25}. An intriguing possibility is that low level bacterial infection coupled with a dysregulated

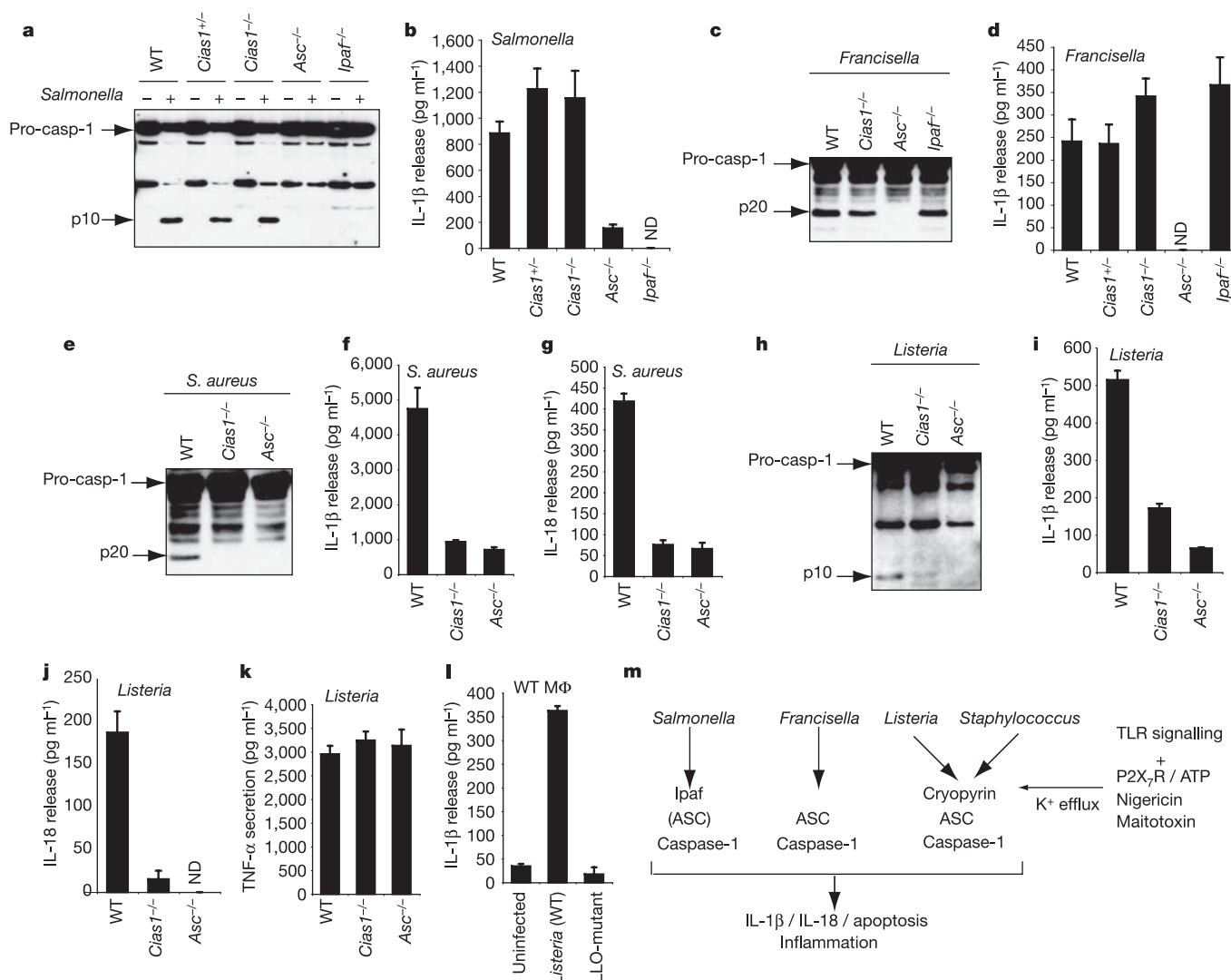


Figure 4 | *L. monocytogenes* and *S. aureus* induce cryopyrin-dependent caspase-1 activation and IL-1 β secretion. a–k, Wild-type, *Cias1*^{-/-}, *Asc*^{-/-} and *Ipaf*^{-/-} macrophages pre-treated with LPS were infected with *S. typhimurium* (a, b), *F. tularensis* (c, d), *S. aureus* (e–g) or *L. monocytogenes* (h–k) and immunoblotted for the p10 or p20 subunit of caspase-1 (a, c, e, h). IL-1 β (b, d, f, i), IL-18 (g, j) or TNF (k) secretion was measured by ELISA. ND, not detected. l, IL-1 β released from wild-type macrophages pre-treated

with LPS and infected with wild-type *L. monocytogenes* or a listeriolysin O (LLO) mutant. All data are representative of 2–4 independent experiments. Bars represent the mean $\pm$ s.d. of triplicate wells. LPS pre-treatment was not essential for IL-1 β release (Supplementary Fig. 6). Cell viability after infection is shown in Supplementary Fig. 4. m, Model for the differential activation of the caspase-1 inflammasome by various pathogens and toxins.

cryopyrin response triggers the symptoms associated with these diseases.

METHODS

Cias1, Nod2, Asc and Ipaf mutant cells and mice. *Asc*^{-/-} and *Ipaf*^{-/-} mice have been described¹. *Cias1*^{-/-} and *Nod2*^{-/-} mice are described in Supplementary Figs 1 and 3. Macrophage and bacterial cultures are described in detail in the Supplementary Methods. Briefly, macrophages primed overnight with 50 ng ml⁻¹ ultra-pure LPS were infected at a multiplicity of infection of 50 (30 for *F. tularensis*) for 1 h (*S. typhimurium*), 2.5 h (*L. monocytogenes*), 3 h (*S. aureus*) or 5 h (*F. tularensis*).

Immunoprecipitation, western blotting and pulse-chase analysis. IL-1 β was immunoprecipitated with goat anti-mouse IL-1 β (clone AF-401-NA; R&D Systems) and blotted with hamster anti-mouse IL-1 β (clone B122; Becton Dickinson). For pulse-chase analyses, macrophages were treated with 1 μ g ml⁻¹ ultra-pure LPS for 3 h and then labelled with 200 μ Ci ml⁻¹ [³⁵S]-methionine (ICN) for 45 min. Labelled cells were treated with 5 mM ATP (Sigma) for 30 min and then placed in fresh medium. Caspase-1 was blotted with rat anti-mouse caspase-1 (clone 4B4; Genentech) and rabbit anti-caspase-1 (sc-514; Santa Cruz Biotechnology). Phospho-I κ B α (Ser32), I κ B α , phospho-ERK and ERK antibodies were from Cell Signalling Technology.

Cytokine ELISAs. TNF and IL-12 p40 secretion were measured by enzyme-linked immunosorbent assay (ELISA; R&D Systems) after 16 h of stimulation with 500 ng ml⁻¹ ultra-pure LPS (List Biologicals), 100 ng ml⁻¹ Pam₃CSK₄ (Invivogen), 10⁸ bacteria per ml HKLM (Invivogen), 5 μ M CpG oligonucleotide (Invivogen ODN1826), 100 ng ml⁻¹ R848 (Invivogen) or 10 μ g ml⁻¹ MDP (Invivogen). The macrophages then were pulsed for 20 min with 5 mM ATP, 20 μ M nigericin (Calbiochem), or 0.5 nM maitotoxin (Dako) and cultured an additional 3 h for IL-1 β and IL-18 secretion. Serum levels of IL-1 β and IL-18 were determined 3 h after intraperitoneal injection with 40 mg kg⁻¹ crude LPS (*Escherichia coli* serotype 0111:B4; Sigma).

Received 7 October; accepted 12 December 2005.

Published online 11 January 2006.

- Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
- Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL- β . *Mol. Cell* **10**, 417–426 (2002).
- Srinivasula, S. M. *et al.* The PYRIN-CARD protein ASC is an activating adaptor for caspase-1. *J. Biol. Chem.* **277**, 21119–21122 (2002).
- Wang, L. *et al.* PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates activation of NF- κ B and caspase-1-dependent cytokine processing. *J. Biol. Chem.* **277**, 29874–29880 (2002).
- Agostini, L. *et al.* NALP3 forms an IL-1 β -processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* **20**, 319–325 (2004).
- Dowds, T. A., Masumoto, J., Zhu, L., Inohara, N. & Nunez, G. Cryopyrin-induced interleukin 1 β secretion in monocytic cells: enhanced activity of disease-associated mutants and requirement for ASC. *J. Biol. Chem.* **279**, 21924–21928 (2004).
- Ting, J. P. & Davis, B. K. CATERPILLER: a novel gene family important in immunity, cell death, and diseases. *Annu. Rev. Immunol.* **23**, 387–414 (2005).
- Perregaux, D. & Gabel, C. A. Interleukin-1 β maturation and release in response to ATP and nigericin. Evidence that potassium depletion mediated by these agents is a necessary and common feature of their activity. *J. Biol. Chem.* **269**, 15195–15203 (1994).
- Solle, M. *et al.* Altered cytokine production in mice lacking P2X₇ receptors. *J. Biol. Chem.* **276**, 125–132 (2001).
- Hogquist, K. A., Nett, M. A., Unanue, E. R. & Chaplin, D. D. Interleukin 1 is processed and released during apoptosis. *Proc. Natl Acad. Sci. USA* **88**, 8485–8489 (1991).
- Yamamoto, M. *et al.* ASC is essential for LPS-induced activation of procaspase-1 independently of TLR-associated signal adaptor molecules. *Genes Cells* **9**, 1055–1067 (2004).
- Poltorak, A. *et al.* Defective LPS signalling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
- Wang, S. *et al.* Murine caspase-11, an ICE-interacting protease, is essential for the activation of ICE. *Cell* **92**, 501–509 (1998).
- Schilling, W. P., Wasylina, T., Dubyak, G. R., Humphreys, B. D. & Sinkins, W. G. Maitotoxin induces biphasic interleukin-1 β secretion and membrane blebbing in murine macrophages. *Mol. Pharmacol.* **66**, 909–920 (2004).
- Li, P. *et al.* Mice deficient in IL-1 β -converting enzyme are defective in production of mature IL-1 β and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).
- Martinon, F., Agostini, L., Meylan, E. & Tschopp, J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. *Curr. Biol.* **14**, 1929–1934 (2004).
- Girardin, S. E. *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J. Biol. Chem.* **278**, 8869–8872 (2003).
- Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
- Maeda, S. *et al.* Nod2 mutation in Crohn's disease potentiates NF- κ B activity and IL-1 β processing. *Science* **307**, 734–738 (2005).
- Damiano, J. S., Newman, R. M. & Reed, J. C. Multiple roles of CLAN (caspase-associated recruitment domain, leucine-rich repeat, and NAIP CIIA HET-E, and TP1-containing protein) in the mammalian innate immune response. *J. Immunol.* **173**, 6338–6345 (2004).
- Hoffman, H. M., Mueller, J. L., Broide, D. H., Wanderer, A. A. & Kolodner, R. D. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome. *Nature Genet.* **29**, 301–305 (2001).
- McDermott, M. F. Genetic clues to understanding periodic fevers, and possible therapies. *Trends Mol. Med.* **8**, 550–554 (2002).
- Feldmann, J. *et al.* Chronic infantile neurological cutaneous and articular syndrome is caused by mutations in *CIA1*, a gene highly expressed in polymorphonuclear cells and chondrocytes. *Am. J. Hum. Genet.* **71**, 198–203 (2002).
- Aganna, E. *et al.* Association of mutations in the NALP3/CIA1/PYPAF1 gene with a broad phenotype including recurrent fever, cold sensitivity, sensorineural deafness, and AA amyloidosis. *Arthritis Rheum.* **46**, 2445–2452 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Dixit laboratory for discussions, and M. Bauer, J. Starks, C. Olsson, M. Osborn, J. Hongo, M. Bever, J. Cupp and H. Maecker for technical assistance. *S. aureus* strains were provided by T. Foster. *L. monocytogenes* LLO mutant was provided by D. Portnoy. This work was supported by NIH grants awarded to D.M.M. and a fellowship from the Giannini Family Foundation awarded to D.S.W.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.M.D. (dixit@gene.com).

Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3

Thirumala-Devi Kanneganti¹, Nesrin Özören¹, Mathilde Body-Malapel¹, Amal Amer¹, Jong-Hwan Park¹, Luigi Franchi¹, Joel Whitfield¹, Winfried Barchet², Marco Colonna², Peter Vandenabeele³, John Bertin^{4†}, Anthony Coyle^{4†}, Ethan P. Grant^{4†}, Shizuo Akira⁵ & Gabriel Núñez¹

Missense mutations in the *CIAS1* gene cause three autoinflammatory disorders: familial cold autoinflammatory syndrome, Muckle-Wells syndrome and neonatal-onset multiple-system inflammatory disease¹. Cryopyrin (also called Nalp3), the product of *CIAS1*, is a member of the NOD-LRR protein family that has been linked to the activation of intracellular host defence signalling pathways^{2,3}. Cryopyrin forms a multi-protein complex termed 'the inflammasome', which contains the apoptosis-associated speck-like protein (ASC) and caspase-1, and promotes caspase-1 activation and processing of pro-interleukin (IL)-1 β (ref. 4). Here we show the effect of cryopyrin deficiency on inflammasome function and immune responses. Cryopyrin and ASC are essential for caspase-1 activation and IL-1 β and IL-18 production in response to bacterial RNA and the imidazoquinoline compounds R837 and R848. In contrast, secretion of tumour-necrosis factor- α and IL-6, as well as activation of NF- κ B and mitogen-activated protein kinases (MAPKs) were unaffected by cryopyrin deficiency. Furthermore, we show that Toll-like receptors and cryopyrin control the secretion of IL-1 β and IL-18 through different intracellular pathways. These results reveal a critical role for cryopyrin in host defence through bacterial RNA-mediated activation of caspase-1, and provide insights regarding the pathogenesis of autoinflammatory syndromes.

To define the role of cryopyrin in inflammatory responses, we generated cryopyrin-deficient mice by homologous recombination using a targeting construct to replace exons I and II of the *cryopyrin* gene (*Cias1*), which encode the pyrin domain of cryopyrin that is essential for effector function of the protein (Supplementary Fig. 1). *Cias1*^{-/-} mice were fertile and appeared healthy when housed in a standard specific pathogen-free environment. We initially investigated the role of cryopyrin in caspase-1-dependent IL-1 β secretion using thioglycollate-elicited peritoneal macrophages and bone marrow-derived macrophages (BMDMs) and multiple bacterial and synthetic ligands. Stimulation of peritoneal macrophages or BMDMs with several TLR2 and TLR4 agonists, including diacylated (Pam₂CGDPPKHPHSF) and triacylated (Pam₃CSK₄) synthetic lipopeptides, lipoteichoic acid, highly purified lipopolysaccharide (LPS) and lipid A induced comparable levels of IL-1 β in wild-type and *Cias1*^{-/-} macrophages (Fig. 1a and Supplementary Fig. 2). Similar results were obtained when macrophages were stimulated with bacterial ligands and treated briefly with ATP (Supplementary Fig. 3), a signal that enhances the secretion of IL-1 β in pre-stimulated macrophages⁵. Incubation of macrophages with muramyl dipeptide (MDP) did not induce secretion of IL-1 β above background levels in

wild-type and *Cias1*^{-/-} macrophages, even after addition of ATP (Fig. 1a; see also Supplementary Fig. 3). Furthermore, production of interferon- α induced by several viruses was unimpaired in macrophages and dendritic cells from *Cias1*^{-/-} mice (Supplementary Fig. 4).

Notably, secretion of IL-1 β and IL-18 induced by the low molecular weight imidazoquinoline compounds imiquimod (R837) and resiquimod (R848), which are known to activate pro-inflammatory responses in the mouse through TLR7 (refs 6, 7), was abrogated in both peritoneal macrophages and BMDMs from *Cias1*^{-/-} mice (Fig. 1a–c). In contrast, cryopyrin was dispensable for the production of the pro-inflammatory cytokines tumour-necrosis factor- α (TNF- α) and IL-6 induced by stimulation with R837 (Fig. 1c). These results indicate that cryopyrin is specifically required for the secretion of IL-1 β and IL-18 induced by the synthetic molecules R837 and R848.

The induction of IL-1 β secretion is thought to involve the upregulation of pro-IL-1 β through transcriptional mechanisms via NF- κ B, followed by a second stimulus that leads to the activation of caspase-1, processing of pro-IL-1 β and release of mature IL-1 β (refs 5, 8). We found that stimulation with R837 induced comparable levels of NF- κ B, extracellular signal-regulated kinase (ERK) and p38 activation in wild-type and *Cias1*^{-/-} macrophages (Fig. 2a). In contrast, activation of NF- κ B and MAPKs was abolished in TLR7- and MyD88-deficient macrophages (Fig. 2b; see also Supplementary Fig. 5), consistent with previous results⁶. Importantly, proteolytic processing of pro-caspase-1 was induced in wild-type macrophages by both R837 and R848, as determined by detection of the mature 20-kDa subunit of caspase-1 (Fig. 2c). Such activation of caspase-1 was abrogated in macrophages lacking cryopyrin (Fig. 2c) or ASC (Fig. 2e), an adaptor that links cryopyrin to caspase-1 (refs 4, 9). In contrast, activation of caspase-1 was unimpaired in *Cias1*^{-/-} macrophages in response to LPS and lipid A (Fig. 2d), further indicating that the ligand-recognition function of cryopyrin is highly specific. In our hands, MDP did not induce proteolytic processing of pro-caspase-1 in mouse macrophages (Fig. 2d), consistent with its inability to induce IL-1 β secretion (Fig. 1a). Notably, activation of caspase-1 induced by R837 proceeded normally in TLR7- or MyD88-deficient macrophages (Fig. 2f, g). These results demonstrate that cryopyrin is essential for caspase-1 processing, independent of NF- κ B and MAPK activation in response to R837 and R848. Furthermore, TLR7 and MyD88 are required for NF- κ B and MAPK activation but are dispensable for caspase-1 activation.

Structurally, R837 and R848 resemble purine bases¹⁰, suggesting

¹University of Michigan Medical School, Department of Pathology and Comprehensive Cancer Center, Ann Arbor, Michigan 48109, USA. ²Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA. ³Department of Molecular Biomedical Research, Ghent University, Zwijnaarde-Gent B-9052, Belgium. ⁴Millennium Pharmaceuticals Inc., Cambridge, Massachusetts 02139, USA. ⁵Department of Host Defense, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. [†]Present addresses: Synta Pharmaceuticals, 45 Hartwell Avenue, Lexington, Massachusetts 02421, USA (J.B. and E.P.G.); MedImmune Inc., 1 MedImmune Way, Gaithersburg, Maryland 20878, USA (A.C.).

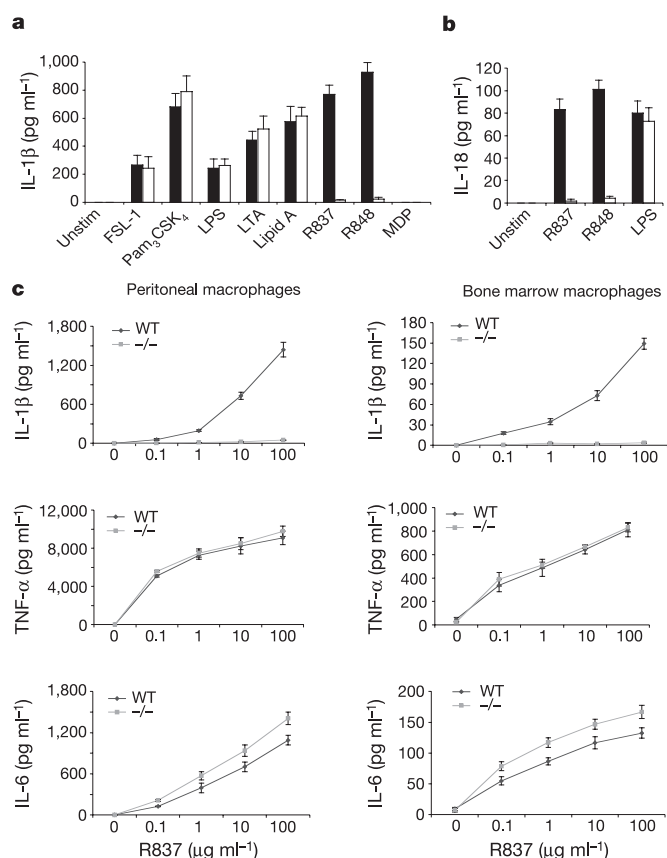


Figure 1 | Cryopyrin is required for IL-1 β and IL-18 secretion in response to imidazoquinoline compounds R837 and R848. **a, b**, Peritoneal macrophages (**a**) and BMDMs (**b**) from wild-type (black bars) or *Cias1*^{-/-} mice (white bars) were stimulated as indicated for 24 h, and cell-free supernatants were analysed by ELISA. LA, lipid A; LTA, lipoteichoic acid; FSL-1, Pam₂CGKPHSE. **c**, Peritoneal macrophages (left panels) and BMDMs (right panels) were stimulated with the indicated concentrations of R837 for 24 h, and cell-free supernatants were analysed by ELISA for production of IL-1 β (top), TNF- α (middle) and IL-6 (bottom). WT, wild type; *-/-*, *Cias1*^{-/-}. Error bars represent the standard deviation of triplicate cultures. Results are representative of at least three separate experiments.

that the natural ligand of cryopyrin could be DNA or RNA. Secretion of IL-1 β and IL-18 was induced by *Escherichia coli* RNA in wild-type and *Cias1*^{+/-} macrophages, but was abolished in *Cias1*^{-/-} macrophages (Fig. 3a and Supplementary Fig. 6). Treatment with chloroquine did not affect the production of IL-1 β induced by bacterial RNA (Supplementary Fig. 7). *E. coli* RNA induced rapid activation of caspase-1 in wild-type but not in *Cias1*^{-/-} macrophages (Fig. 3b). As we found with *E. coli* RNA, stimulation with total RNA from two different bacteria (*Listeria monocytogenes* and *Legionella pneumophila*), but not total RNA from mouse liver, induced activation of caspase-1 in wild-type but not in *Cias1*^{-/-} macrophages (Fig. 3c, d). Treatment of the RNA preparations with RNase abolished their ability to induce caspase-1 activation (Fig. 3d), indicating that RNA, but not a contaminating product, triggers caspase-1 activation. Furthermore, secretion of IL-1 β and processing of pro-caspase-1 induced by bacterial RNA was abrogated in macrophages lacking ASC, but was unimpaired in TLR7- or MyD88-deficient macrophages (Fig. 3e and Supplementary Fig. 8). These results show that bacterial RNA activates caspase-1 and IL-1 β secretion, and that these events are mediated through cryopyrin and ASC, and are independent of TLR7 and MyD88.

Mononuclear cells from patients with autoinflammatory syndromes spontaneously secrete IL-1 β and IL-18, and show enhanced

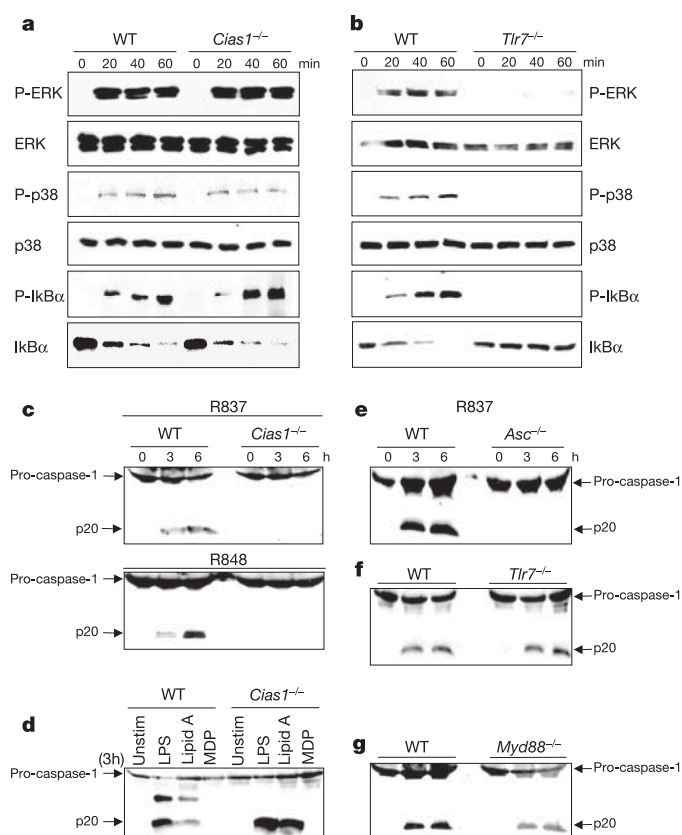


Figure 2 | Caspase-1 processing and NF- κ B activation in mutant macrophages stimulated with R837 or R848. **a, b**, BMDMs from wild-type and *Cias1*^{-/-} mice (**a**) or wild-type and *Tlr7*^{-/-} mice (**b**) were stimulated with R837 (5 μ g ml⁻¹) for the indicated times, and extracts were immunoblotted with antibodies that recognize phosphorylated (P)-ERK, p38, IκB α and total protein. IκB α phosphorylation can be used as a measure of NF- κ B activation. **c–g**, BMDMs from wild-type and the indicated mutant mice were stimulated with R837 (5 μ g ml⁻¹) or R848 (5 μ g ml⁻¹) (**c, e–g**), LPS, lipid A (LA) or MDP (10 μ g ml⁻¹) (**d**). Cell extracts were immunoblotted with an antibody against caspase-1. Arrows denote pro-caspase-1 and its processed p20 subunit.

production of IL-1 β and IL-18 in response to low amounts of LPS^{4,11}. This is consistent with the observation that disease-associated cryopyrin mutations show constitutive activity⁹. We therefore examined the ability of low doses of LPS to cooperate with low amounts of various microbial ligands in the production of IL-1 β . LPS synergized with R837, but not with lipopeptides, lipoteichoic acid, lipid A, flagellin or CpG oligodeoxynucleotide for the secretion of IL-1 β (Fig. 4a). Such enhancement of LPS-mediated IL-1 β production by R837 was abrogated in *Cias1*^{-/-} macrophages (Fig. 4a). The susceptibility to high doses of highly purified LPS induced comparable lethality in wild-type and *Cias1*^{-/-} mice (Supplementary Fig. 9). The synergy between LPS and R837 might be explained, at least in part, by the ability of LPS to induce cryopyrin expression¹². Consistent with our *in vitro* results, co-administration of R837 and LPS to mice induced higher levels of IL-1 β in serum than injection of R837 or LPS alone (Fig. 4b). Furthermore, the enhancement of the LPS response by R837 was abolished in *Cias1*^{-/-} mice (Fig. 4b). The serum levels of IL-6 and TNF- α were also enhanced by co-injection of LPS and R837 when compared to those observed after administration of each molecule alone (Fig. 4c, d). Moreover, reduced levels of IL-6 and TNF- α were detected in the serum of *Cias1*^{-/-} mice after stimulation with LPS plus R837 (Fig. 4c, d), presumably owing to the induction of IL-6 and TNF- α by IL-1 β and/or IL-18 *in vivo*¹³.

These studies show that cryopyrin has an essential role in the secretion of IL-1 β and IL-18 by controlling the activation of

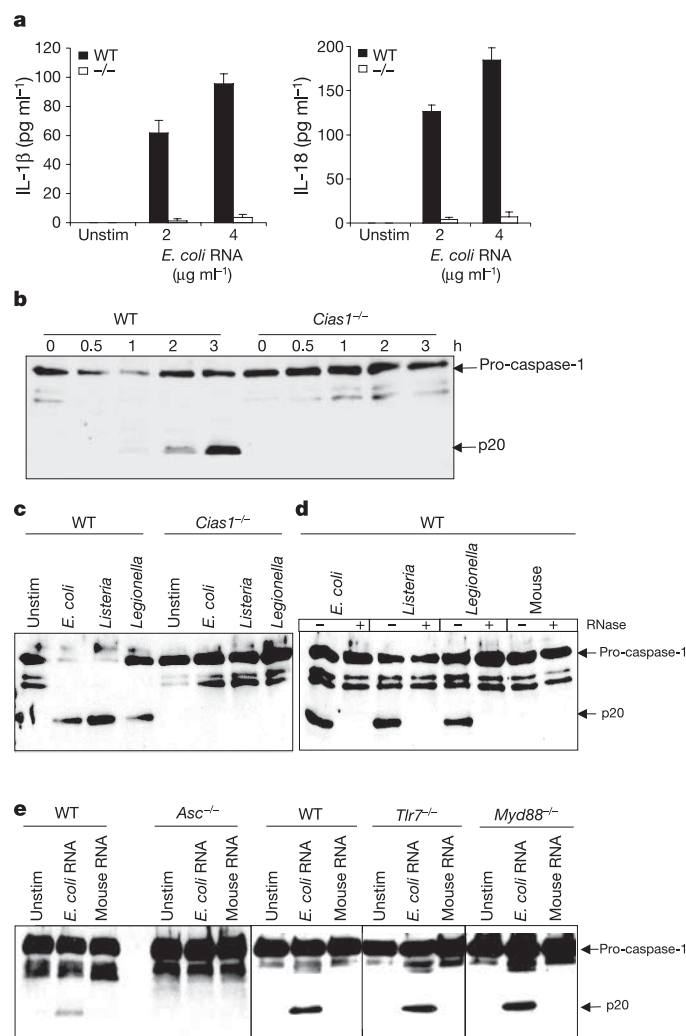


Figure 3 | Cryopyrin is essential for activation of caspase-1 in response to bacterial RNA. **a**, BMDMs from wild-type (black bars) or *Cias1*^{-/-} mice (white bars) were stimulated as indicated, and cell-free supernatants were analysed by ELISA for production of IL-1 β or IL-18. **b–d**, BMDMs from wild-type and *Cias1*^{-/-} mice were stimulated with RNA purified from *E. coli* (**b**) or the indicated bacteria (**c**) with (+) or without (–) RNase digestion (**d**), and cell extracts were immunoblotted with an antibody against caspase-1. **e**, BMDMs from wild-type and *Asc*^{-/-}, *Tlr7*^{-/-} and *Myd88*^{-/-} mice were stimulated with purified RNA from *E. coli* or mouse liver for 3 h.

pro-caspase-1 in response to bacterial RNA or the synthetic compounds R837 and R848. We provide evidence that secretion of IL-1 β and IL-18 is regulated by separate signalling pathways that are independently controlled by TLR and NOD-LRR proteins. Previous studies, using an overexpression system in human HEK293 cells, have suggested that MDP mediates cryopyrin-mediated activation of caspase-1 and maturation of pro-IL-1 β (ref. 14). We found no evidence that MDP induces the activation of caspase-1 or the release of IL-1 β in mouse macrophages, consistent with previous results¹⁵. Our studies, however, do not rule out the possibility that cryopyrin responds to MDP under certain conditions in the mouse and/or human system.

Our results indicate that TLRs and cryopyrin are involved in the activation of immune responses induced by the same or similar microbial structures^{6,16–18}. Bacterial RNA might be derived from phagocytosed bacteria or lysis of bacteria in the extracellular space and transported into the host cytosol, leading to cryopyrin recognition and activation. The mechanism by which cryopyrin and

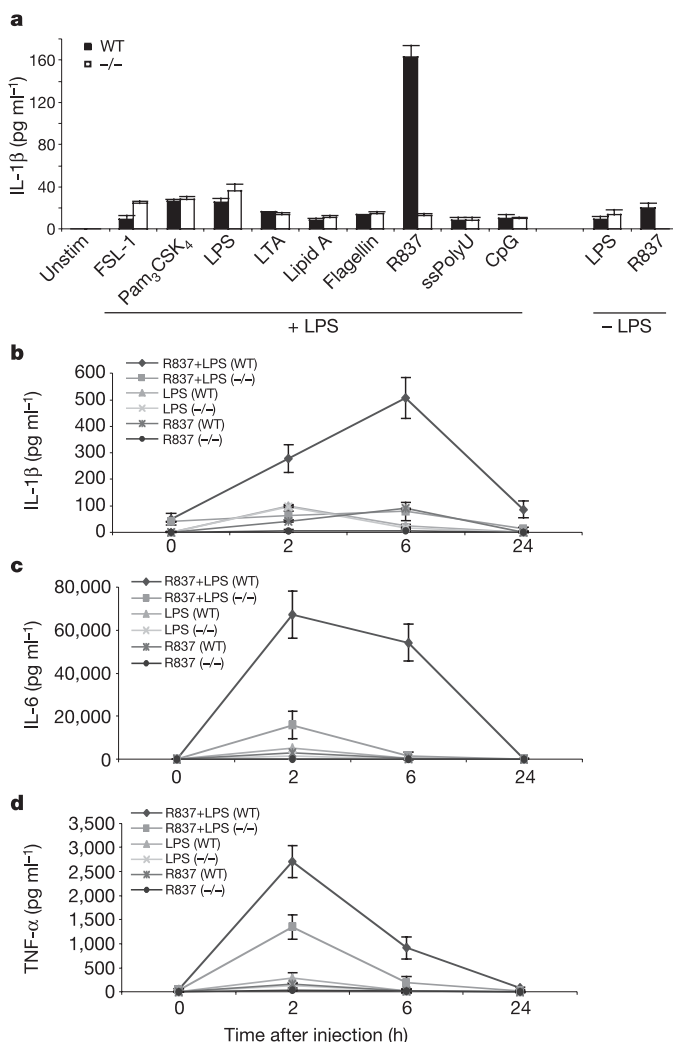


Figure 4 | R837 and LPS cooperate in the production of pro-inflammatory cytokines in a cryopyrin-dependent manner. **a**, BMDMs from wild-type (WT) or *Cias1*^{-/-} mice (–/–) were co-stimulated with LPS and the indicated stimuli for 24 h or left unstimulated (unstim). Cell-free supernatants were analysed by ELISA for IL-1 β production. **b–d**, Groups of wild-type and *Cias1*^{-/-} mice ($n = 7$ mice per group) were co-injected with LPS (200 μ g), R837 (200 μ g) or LPS plus R837 (200 μ g each), and levels of IL-1 β (**b**), IL-6 (**c**) or TNF- α (**d**) in serum were determined by ELISA at the indicated times. Error bars represent standard deviation of serum values.

TLR7/TLR8 discriminate between microbial and endogenous RNA remains poorly understood, but differences in nucleoside modification and the polyA-tail between bacterial and mammalian RNA might be involved^{19,20}. R837 and R848 are used as immune response modifiers in the clinic and are being considered as vaccine adjuvants^{21–24}. The potent adjuvant, antiviral and antitumour activity exerted by the imidazoquinoline compounds could be explained by their ability to activate both TLR and cryopyrin signalling. The identification of cryopyrin as a critical factor for caspase-1 activation induced by bacterial RNA has implications for host defence and RNA-based vaccines as well as for our understanding of inflammatory diseases.

METHODS

Mice. *Cias1*-knockout mice were generated by homologous recombination in embryonic stem cells by replacing exons I and II of the *cryopyrin/Cias1* gene (encoding the amino-terminal Pyrin domain) with an IRES- β -gal-neomycin-resistance cassette using a targeting vector (Supplementary Fig. 1). A positive embryonic stem cell clone was used to generate chimaeric mice. 129/C57BL/6

chimaeric mice were crossed with C57BL/6 females to generate heterozygous mice. *Cias1*-knockout and wild-type mice were generated by crossing male and female heterozygous mice. *Tlr7*, *Myd88* and *Asc* knockout mice have been described^{6,25}.

Microbial ligands and antibodies. Ultrapure LPS from *E. coli* 0111:B4 (Invivo-gen) was used in all experiments. FSL-1, Pam₃CSK₄, lipid A, flagellin, R837 and CpG oligonucleotide were purchased from InvivoGen and muramyl dipeptide was purchased from Bachem. Purified total RNA from *E. coli* and mouse liver was purchased from Ambion. Total RNA from *L. pneumophila* and *L. monocytogenes* was prepared using a RiboPure-bacteria kit (Ambion) and RNase (Novagen). The 260/280 absorbance of RNA was 1.7/2. Rabbit anti-mouse caspase-1 and anti-cryopyrin antibodies have been previously described^{12,26}. Antibodies against mouse Ik- β , phospho-Ik- β , p38 and phospho-p38 were from Cell Signalling. **Western blotting.** For analysis of caspase-1 activation, macrophages were cultured with ligands for 1–3 h and then with medium containing 5 mM ATP (Sigma) for 30 min. Extracts were prepared, transferred to nitrocellulose membranes and immunoblotted with primary antibodies, and then proteins were detected by enhanced chemiluminescence as previously described²⁷.

Measurements of cytokines. BMDMs and peritoneal macrophages were prepared as described²⁵. Cells were stimulated with various microbial and synthetic ligands for 24 h, and supernatants were analysed for IL-1 β , IL-18, TNF- α and IL-6 secretion. FSL-1, Pam₃CSK₄, LPS, lipoteichoic acid, lipid A and MDP were used at 1 μ g ml⁻¹. For analysis of the cooperation between LPS and various ligands, 10 ng ml⁻¹ LPS and 10 ng ml⁻¹ of ligand were used. Mouse cytokines were measured in culture supernatants using enzyme-linked immunoabsorbent assay (ELISA) kits from R&D Systems.

Received 9 October; accepted 12 December 2005.

Published online 11 January 2006.

- Stojanov, S. & Kastner, D. L. Familial autoinflammatory diseases: genetics, pathogenesis and treatment. *Curr. Opin. Rheumatol.* **17**, 586–599 (2005).
- Inohara, N., Chamaillard, M., McDonald, C. & Nunez, G. NOD-LRR proteins: role in host-microbial interactions and inflammatory disease. *Annu. Rev. Biochem.* **74**, 355–383 (2005).
- Athman, R. & Philpott, D. Innate immunity via Toll-like receptors and Nod proteins. *Curr. Opin. Microbiol.* **7**, 25–32 (2004).
- Agostini, L. et al. NALP3 forms an IL-1 β -processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* **20**, 319–325 (2004).
- Perregaux, D. G., McNiff, P., Laliberte, R., Conklyn, M. & Gabel, C. A. ATP acts as an agonist to promote stimulus-induced secretion of IL-1 β and IL-18 in human blood. *J. Immunol.* **165**, 4615–4623 (2000).
- Hemmi, H. et al. Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nature Immunol.* **3**, 196–200 (2002).
- Jurk, M. et al. Human TLR7 or TLR8 independently confer responsiveness to the antiviral compound R-848. *Nature Immunol.* **3**, 499 (2002).
- Dinarello, C. A. Interleukin-1 β , interleukin-18, and the interleukin-1 β converting enzyme. *Ann. NY Acad. Sci.* **856**, 1–11 (1998).
- Dowds, T. A., Masumoto, J., Zhu, L., Inohara, N. & Nunez, G. Cryopyrin-induced interleukin 1 β secretion in monocytic cells: enhanced activity of disease-associated mutants and requirement for ASC. *J. Biol. Chem.* **279**, 21924–21928 (2004).
- Dockrell, D. H. & Kinghorn, G. R. Imiquimod and resiquimod as novel immunomodulators. *J. Antimicrob. Chemother.* **48**, 751–755 (2001).
- Janssen, R., Verhard, E., Lankester, A., Ten Cate, R. & van Dissel, J. T. Enhanced interleukin-1 β and interleukin-18 release in a patient with chronic infantile neurologic, cutaneous, articular syndrome. *Arthritis Rheum.* **50**, 3329–3333 (2004).
- O'Connor, W. Jr, Harton, J. A., Zhu, X., Linhoff, M. W. & Ting, J. P. Cutting edge: CIAS1/cryopyrin/PYPAF1/NALP3/CATERPILLER 1.1 is an inducible inflammatory mediator with NF- κ B suppressive properties. *J. Immunol.* **171**, 6329–6333 (2003).
- Netea, M. G., Kullberg, B. J., Verschuere, I. & Van Der Meer, J. W. Interleukin-18 induces production of proinflammatory cytokines in mice: no intermediate role for the cytokines of the tumour necrosis factor family and interleukin-1 β . *Eur. J. Immunol.* **30**, 3057–3060 (2000).
- Martinon, F., Agostini, L., Meylan, E. & Tschopp, J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. *Curr. Biol.* **14**, 1929–1934 (2004).
- Kobayashi, K. S. et al. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
- Heil, F. et al. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526–1529 (2004).
- Lund, J. M. et al. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **101**, 5598–5603 (2004).
- Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).
- Kariko, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175 (2005).
- Koski, G. K. et al. Cutting edge: innate immune system discriminates between RNA containing bacterial versus eukaryotic structural features that prime for high-level IL-12 secretion by dendritic cells. *J. Immunol.* **172**, 3989–3993 (2004).
- Craft, N. et al. The TLR7 agonist imiquimod enhances the anti-melanoma effects of a recombinant *Listeria monocytogenes* vaccine. *J. Immunol.* **175**, 1983–1990 (2005).
- Harrison, C. J., Jenks, L., Voychevski, T. & Bernstein, D. I. Modification of immunological responses and clinical disease during topical R-837 treatment of genital HSV-2 infection. *Antiviral Res.* **10**, 209–223 (1988).
- Tomai, M. A. et al. Immunomodulating and antiviral activities of the imidazoquinoline S-28463. *Antiviral Res.* **28**, 253–264 (1995).
- Marks, R. et al. Imiquimod 5% cream in the treatment of superficial basal cell carcinoma: results of a multicenter 6-week dose-response trial. *J. Am. Acad. Dermatol.* **44**, 807–813 (2001).
- Özören, N. et al. Distinct roles of TLR2 and the adaptor ASC in IL-1 β /IL-18 secretion and host defense against *Listeria monocytogenes*. *J. Immunol.* (submitted).
- Lamkanfi, M., Kalai, M., Saelens, X., Declercq, W. & Vandenabeele, P. Caspase-1 activates nuclear factor of the κ -enhancer in B cells independently of its enzymatic activity. *J. Biol. Chem.* **279**, 24785–24793 (2004).
- Mariathasan, S. et al. Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Golenbock and P. Lin for providing mouse bone marrow, S. Bauer for the gift of R848, and J. Ting for anti-cryopyrin antibody. We thank C. McDonald and N. Inohara for discussions and advice, and the Cellular Immunology Core Facility of the University of Michigan Cancer Center for technical support. This work was supported by grants from the National Institutes of Health to G.N. T.-D.K. is supported by an NIH training grant. L.F. is the recipient of a postdoctoral fellowship from the Fondazione Italiana Ricerca sul Cancro. M.B.-M. is supported by a postdoctoral fellowship from the Fondation pour la Recherche Médicale.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.N. (bcx@med.umich.edu).

Gout-associated uric acid crystals activate the NALP3 inflammasome

Fabio Martinon¹, Virginie Pétrilli¹, Annick Mayor¹, Aubry Tardivel¹ & Jürg Tschopp¹

Development of the acute and chronic inflammatory responses known as gout and pseudogout are associated with the deposition of monosodium urate (MSU) or calcium pyrophosphate dihydrate (CPPD) crystals, respectively, in joints and periarticular tissues. Although MSU crystals were first identified as the aetiological agent of gout in the eighteenth century¹ and more recently as a 'danger signal' released from dying cells², little is known about the molecular mechanisms underlying MSU- or CPPD-induced inflammation. Here we show that MSU and CPPD engage the caspase-1-activating NALP3 (also called cryopyrin) inflammasome, resulting in the production of active interleukin (IL)-1 β and IL-18. Macrophages from mice deficient in various components of the inflammasome such as caspase-1, ASC and NALP3 are defective in crystal-induced IL-1 β activation. Moreover, an impaired neutrophil influx is found in an *in vivo* model of crystal-induced peritonitis in inflammasome-deficient mice or mice deficient in the IL-1 β receptor (IL-1R). These findings provide insight into the molecular processes underlying the inflammatory conditions of gout and pseudogout, and further support a pivotal role of the inflammasome in several auto-inflammatory diseases.

The notion of autoinflammatory diseases delineates a heterogeneous group of pathologies characterized by spontaneous periodic inflammation and fever in the absence of infectious or autoimmune causes³. Hereditary periodic fevers, systemic onset juvenile idiopathic arthritis, Still's disease, Behçet's disease and the metabolic disorders gout and pseudogout are examples of such inflammatory maladies. Increased production of the inflammatory cytokine IL-1 β was recently identified as the cause of several autoinflammatory diseases, providing clear evidence for a pivotal role of this cytokine in triggering auto-inflammation^{4–8}. IL-1 β , also known as the endogenous pyrogen, is a highly inflammatory cytokine whose production is tightly controlled by at least three distinct steps⁹. The first step involves the production of the pro-IL-1 β protein (p35); this is followed by cleavage of the precursor pro-IL-1 β to produce the active IL-1 β protein (p17), and finally IL-1 β is released into the extracellular environment. The middle step, processing of pro-IL-1 β , involves the activation of a caspase-1-activating complex, the best characterized being the inflammasome^{10,11}.

Upon activation, the inflammasome is formed by a member of the NALP protein family, such as NALP1, NALP2 or NALP3, and the adaptor protein ASC that connects the NALPs with caspase-1 (ref. 12). Signals and mechanisms leading to inflammasome activation are still poorly understood. Muramyl dipeptide (MDP), a degradation product of the bacterial cell wall component peptidoglycan and contaminant of crude lipopolysaccharide (LPS), was recently shown to activate a NALP3 inflammasome¹³ through the leucine-rich repeat domain of NALP3, suggesting that NALPs, like Toll-like receptors (TLRs), are fundamental for microbial

detection¹⁴. However, the inflammasome is also proficient in sensing stress or endogenous danger signals, such as extracellular ATP or hypotonic stress^{10,11,15}. Recently, MSU crystals were identified as a danger signal formed after release of uric acid from dying cells². This observation, and the well-known role of uric acid crystals in gouty arthritis¹⁶, prompted us to investigate whether MSU crystals could activate the inflammasome.

Cells from the differentiated monocytic cell line THP1 were incubated with MSU crystals. Maturation of IL-1 β was indeed detected after stimulation with as little as 10 $\mu\text{g ml}^{-1}$ of the crystals (Fig. 1a). The caspase-1 dependency of the pro-IL-1 β cleavage was confirmed by addition of the caspase-1 inhibitor zYVAD-fmk, which completely blocked MSU-induced IL-1 β activation (Fig. 1a). CPPD, another type of pathogenic crystal involved in calcium pyrophosphate deposition disease, also known as pseudogout, was as active as MSU (Fig. 1b). Crystal-induced IL-1 β processing was specific for pathogenic agents, as the non-inflammatory allopurinol or diamond crystals and particulate elements such as zymosan and aluminium powder failed to induce pro-IL-1 β processing (Fig. 1c), despite their similar size and/or chemical composition. Compared to the known activators of the inflammasome (that is, crude LPS, ATP), MSU and CPPD were more active^{11,13} (Fig. 1c). This superior potency was particularly evident when analysing processing of pro-IL-18, the second known substrate of caspase-1 (Fig. 1c). Previously, we demonstrated that the inflammatory caspases are cleaved and released along with active IL-1 β after activation of the inflammasome¹³. This was also observed when cells were treated with MSU and CPPD (Fig. 1c, d). In order to exclude the possibility that crystal-mediated activation of caspase-1 is a unique property of the THP1 cell line only, MSU and CPPD were added to purified human monocytes. As shown in Fig. 1d, a strong response to both pathogenic crystals was also elicited in primary cells.

In order to provide direct evidence for the involvement of the inflammasome in crystal-induced inflammation, we analysed peritoneal macrophages (PM Φ s) derived from mice deficient in various key proteins of the inflammasome complex or other proinflammatory pathways. Given the absence and/or rapid degradation of pro-IL-1 β in PM Φ s *ex vivo*, and because we failed to see any direct induction of the transcription or translation of pro-IL-1 β by MSU or CPPD, we stimulated TLR4 in PM Φ s with highly purified LPS to induce the synthesis of the cytokine^{11,13}. Consistent with our previous findings in human monocytes, murine PM Φ s stimulated with MSU or CPPD activated caspase-1 and secreted mature IL-1 β (Fig. 2a). Maturation was abolished in PM Φ s from caspase-1-deficient mice, confirming the specificity of the activation. As expected, MyD88-deficient PM Φ s did not produce mature IL-1 β due to their defective TLR signalling, resulting in a failure to produce pro-IL-1 β after LPS pre-stimulation (Fig. 2a). Nevertheless, MyD88^{-/-} PM Φ s still activated caspase-1 (Fig. 2a), further suggesting that this activation is

¹Department of Biochemistry, University of Lausanne, Chemin des Boveresses 155, 1066 Epalinges, Switzerland.

TLR independent and is consistent with a possible involvement of the inflammasome^{13,15}. ASC is a crucial adaptor protein required for the recruitment of caspase-1 to the NALP platform of inflammasomes¹². ASC-deficient PM Φ s did not produce any mature IL-1 β after stimulation by MSU and CPPD crystals (Fig. 2b).

The human genome harbours a repertoire of 14 NALPs. It is currently not clear how many of them form inflammasomes. NALP3 is expressed in both monocytes and macrophages and is well conserved in human and mouse. Its ability to form an inflammasome and to drive inflammation in humans is well supported by its implication in many hereditary autoinflammatory syndromes⁴. We considered that the NALP3 inflammasome was possibly implicated in crystal-induced caspase-1 activation and we therefore generated NALP3-deficient mice (Supplementary Fig. 1 and V.P., F.M. and J.T., manuscript in preparation). Similar to PM Φ s from *Asc*^{-/-} mice, IL-1 β release was impaired in NALP3-deficient PM Φ s upon MSU and CPPD exposure (Fig. 2c). IL-1 β induction by ATP, the other known non-microbial stimulus of inflammasomes, was also dependent on NALP3 (Fig. 2c). Whereas blocking of the ATP receptor P2X₇ inhibited ATP-driven inflammasome activation, it had no effect on MSU-induced activation, indicating that the two inflammasome-activating pathways act independently (Supplementary Fig. 2).

In addition to cytokines whose activity is dependent on caspase-1 activation, MSU and CPPD are known to induce release of other cytokines such as TNF^{17,18}, suggesting additional, inflammasome-

independent activities of the crystals. When assaying the release of TNF, we realized that the production of TNF was relatively slow and was preceded by the release of IL-1 β ¹⁹ (Fig. 3a). It was therefore possible that TNF secretion was initiated, at least in part, by the released mature IL-1 β . Indeed, blocking the maturation of IL-1 β with zYVAD-fmk considerably reduced the production of TNF induced by MSU and CPPD, without affecting TNF production by the TLR2 agonist zymosan (Fig. 3a). Similarly, IL-1ra, a natural inhibitor of IL-1 signalling, significantly affected the production of TNF and IL-6 by human monocytes (Fig. 3b). These results suggest that the processing of IL-1 β is a proximal event in the inflammatory cascade initiated by pathogenic crystals, possibly explaining the extraordinary success of IL-1ra in the treatment of some auto-inflammatory diseases^{20,21}.

Colchicine is another drug that is frequently used for the treatment of autoinflammatory diseases, including familial Mediterranean fever, acute gout and pseudogout episodes²². Pre-treatment with intravenous colchicine before intra-articular MSU injections greatly reduces inflammation²³, suggesting that colchicine targets the initial phase of inflammation. We therefore investigated the role of colchicine in crystal-induced maturation of IL-1 β . As shown in Fig. 3c, pre-treatment with colchicine, but not its solvent ethanol, completely blocked the processing of IL-1 β . In contrast, colchicine did not affect IL-1 β activation by extracellular ATP, indicating that the drug acts upstream of inflammasome activation. Taken together, the above

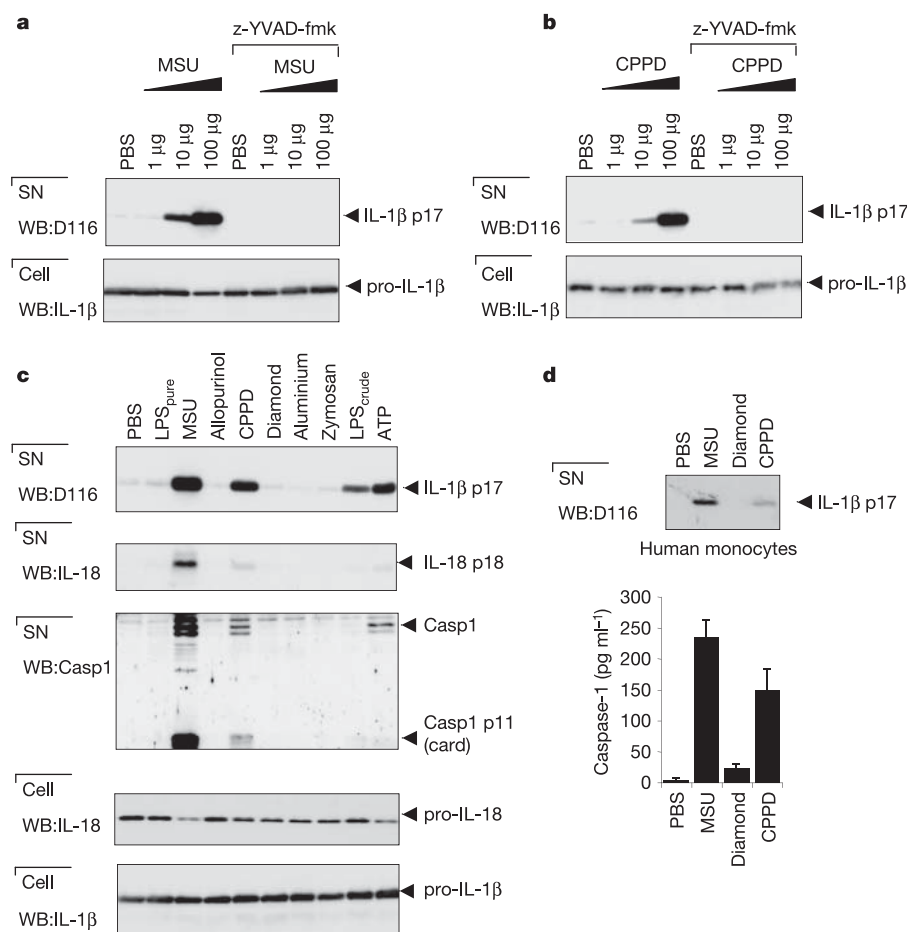


Figure 1 | Monosodium urate (MSU) and calcium pyrophosphate dihydrate (CPPD) crystals activate IL-1 β cleavage and release. **a–c**, THP1 cells were stimulated for 6 h with the indicated amounts (per ml) of MSU crystals (**a**), CPPD crystals (**b**) or with 50 μ g ml⁻¹ of pure LPS, MSU crystals, allopurinol crystals, CPPD crystals, diamond crystals, aluminium particles, zymosan, crude preparations of LPS, or 5 mM of extracellular ATP as indicated (**c**).

Supernatants (SN) were analysed for the presence of mature IL-1 β , IL-18 or caspase-1, and cell extracts (Cell) for the presence of pro-IL-1 β and pro-IL-18. WB, western blot. **d**, Human monocytes were stimulated with 50 μ g ml⁻¹ of the indicated crystals for 6 h and analysed by western blot for IL-1 β activation or by ELISA for released caspase-1 and IL-1 β . Values are $\pm$ s.e.m.

results indicate that crystals are proinflammatory by virtue of their capacity to activate the NALP3 inflammasome.

Clinically, gout and pseudogout are associated with oedema and erythema of the joints, with consequent severe pain, conditions that are associated with strong infiltration of neutrophils in the intra-articular and periarticular spaces. This marked neutrophil influx can be reproduced experimentally in mice by intraperitoneal injection of crystals²⁴. We used this well-established model to investigate the *in vivo* role of the inflammasome in crystal-induced inflammation. MSU, CPPD or allopurinol crystals were injected and the peritoneal recruitment of neutrophils was analysed 6 h later. Both MSU and CPPD crystals elicited a considerable increase in the recruitment of neutrophils compared with PBS or allopurinol when injected in wild-type C57BL/6 mice (Fig. 4a). Although neutrophil influx was slightly increased in BALB/c mice, both BALB/c and C57BL/6 strains adequately reproduced crystal-induced inflammation (Fig. 4 and data not shown). Importantly, when pathogenic crystals were injected in mice deficient in caspase-1 or ASC, neutrophil influx was markedly impaired (Fig. 4b, c), indicating a pivotal role of the inflammasome and IL-1 β in this process. In agreement with this notion was the observation that IL-1R-deficient mice exhibited a similarly reduced recruitment of neutrophils after MSU and CPPD injection (Fig. 4d). In contrast, zymosan-induced neutrophil influx

was not affected by ASC or IL-1R deficiency.

Gout and pseudogout are two common causes of inflammatory joint diseases. Despite differences underlying their pathogenesis, their clinical presentation and treatment share many common features. On the basis of our findings that pathogenic crystal-mediated IL-1 β maturation requires the inflammasome components NALP3, ASC and caspase-1, we propose that both aetiological agents of gout and pseudogout (that is, MSU and CPPD) mediate inflammation in an inflammasome-dependent manner. This notion is further supported by clinical data demonstrating that colchicine, a drug able to resolve the initial inflammatory phase of both gout and pseudogout, blocks IL-1 β maturation by MSU and CPPD. Combining colchicine's known mode of action as an inhibitor of microtubule assembly with the observation that colchicine blocks crystal-induced IL-1 β generation upstream of inflammasome activation (Fig. 3c), it is likely that the drug acts at the level of crystal endocytosis and/or presentation to the inflammasome²⁵. Notably, MSU uptake was also proposed to be partly dependent on the presence of TLR2 and TLR4 (ref. 26). The mechanism whereby endocytosed MSU and CPPD are sensed by the NALP3 inflammasome is currently not known, nor is it clear whether the crystals directly interact with NALP3 or whether sensing occurs via intermediary protein(s). Because inflammasome-activating MSU crystals and inflammasome neutral allopurinol are chemically and structurally similar and, additionally, are both internalized, inflammasomes must have the capacity to distinguish between subtle differences in crystal surface charge or form.

Microbial components (pathogen-associated molecular patterns

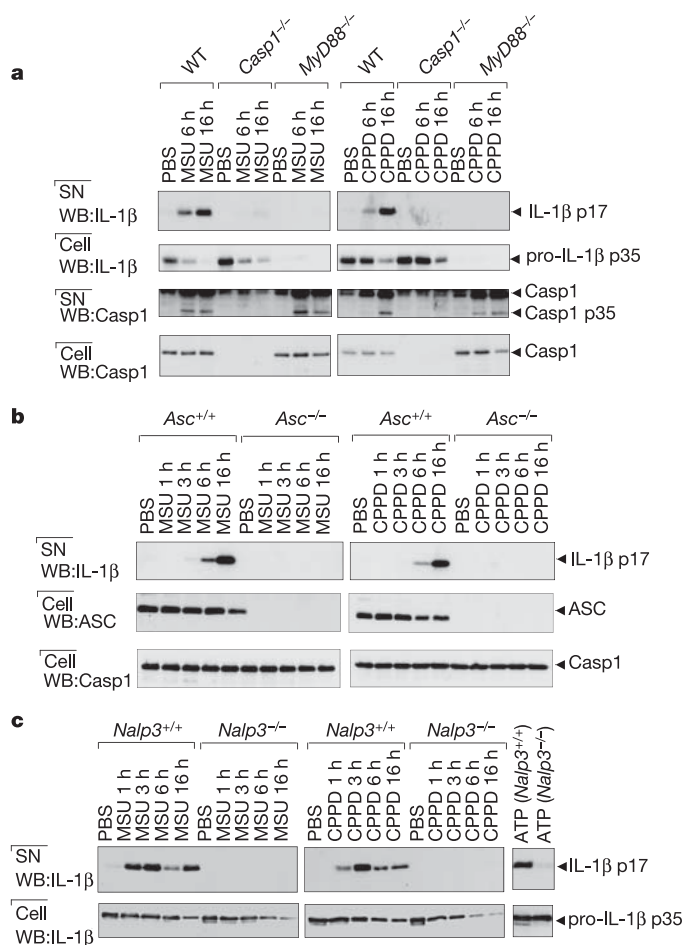


Figure 2 | The NALP3 inflammasome is required for the maturation of IL-1 β . **a–c**, Mouse macrophages from wild-type (WT), caspase-1 (Casp1)- or MyD88-deficient mice (**a**), ASC-deficient mice or littermate controls (**b**), and NALP3-deficient mice or littermate controls (**c**) were stimulated as indicated in the presence of ultra-pure LPS (1 μ g ml⁻¹, Alexis or Invivogen) in order to induce the synthesis of precursor pro-IL-1 β . In **c**, ultra-pure LPS was added 1 h before stimulation. Supernatant (SN) or cell extracts (Cell) were analysed by western blot as indicated.

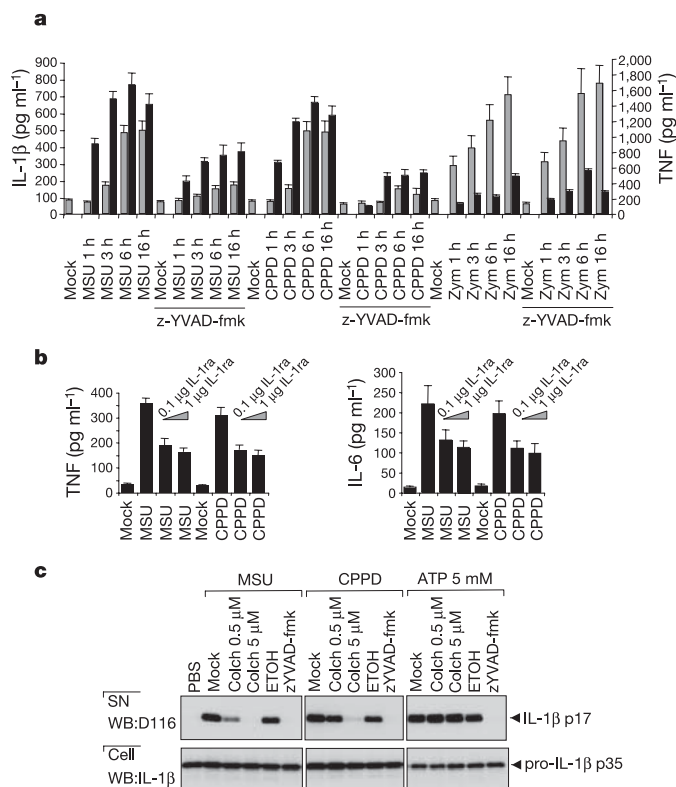


Figure 3 | IL-1 β maturation is an early event after MSU and CPPD stimulation, and is blocked by colchicine. **a**, THP1 cells were stimulated with MSU, CPPD or zymosan (Zym) for the indicated times in the presence or absence of the caspase-1 inhibitor zVAD-fmk. Supernatants were analysed for TNF (grey bars) and IL-1 β (black bars) production by ELISA. Values are $\pm$ s.e.m. **b**, Human monocytes were incubated with MSU or CPPD in the presence of two concentrations of IL-1ra. TNF and IL-6 production was monitored by ELISA. Values are $\pm$ s.e.m. **c**, THP1 cells were stimulated with MSU, CPPD or ATP in the presence or absence of colchicine (Colch). Maturation of IL-1 β was analysed by western blot.

(PAMPs)) provide signals that alert our immune system to danger and promote the innate generation of immunity²⁷. However, PAMPs (non-self) are not the only triggers of innate immunity. Innate immunity is able to recognize abnormal self or danger signals, such as uric acid released by injured cells^{2,28}. How these danger signals are recognized by cells is mostly unknown, but based on our results inflammasomes probably constitute some of the long-sought proximal sensors for stress or danger signals designed to initiate inflammation.

In addition to gouty inflammation, the NALP3 inflammasome is also implicated in other autoinflammatory diseases. Specific gain-of-function mutations in the NALP3 protein lead to three related familial autoinflammatory diseases: Muckle-Wells syndrome, familial cold autoinflammatory syndrome and chronic infantile neurologic cutaneous and articular syndrome^{4,29}. In patients with these diseases, mutations in NALP3 lead to a constitutive processing of IL-1 β ³⁰. In the case of gout and pseudogout, aberrant NALP3 inflammasome activation is not genetic, but mediated by local deposition of crystals. Importantly, inflammation in hereditary periodic fevers patients with mutations in NALP3 can be markedly improved by treatments designated to block IL-1 β ^{20,21}. Owing to the similarity between NALP3-mediated hereditary periodic fevers and gout and pseudogout, we can anticipate that similar treatments could benefit gout and pseudogout patients. It is also reasonable to foresee that further identification of additional inflammasome-activating endogenous danger signals will probably shed some light on the molecular aetiology of other autoinflammatory diseases such as systemic onset juvenile idiopathic arthritis and Behçet's disease

that share similarity with hereditary periodic fevers, gout or pseudogout.

METHODS

Primary human monocyte and THP1 preparation and stimulation. THP1 cells were stimulated for 3 h with 0.5 μ M of PMA the day before stimulation, as described¹⁰. This treatment increases the phagocytic properties of the cells and induces a constitutive production of pro-IL-1 β . Human monocytes were purified as described previously³⁰. All cells were stimulated in OptiMEM medium as indicated. Human mature IL-1 β was detected with a specific antibody directed against the cleaved epitope (D116) from Cell Signaling.

Mouse macrophage preparation. Eight-to-twelve-week-old mice of indicated genotypes were injected intraperitoneally with 4% thioglycollate solution, and macrophages were collected by peritoneal lavage 3 days later. Cells were plated at the density of 7×10^5 cells in 12-well dishes and non-adherent cells were removed after 3 h. Cells were cultured in RPMI complemented with 10% FCS, sodium pyruvate, penicillin/streptomycin and L-glutamine. All cells were stimulated in OptiMEM medium.

In vivo mouse peritonitis model. Peritonitis was induced by injection of 1 mg of crystals or 0.2 mg of zymosan in 0.5-ml sterile PBS. After 6 h, mice were killed by CO₂ exposure and peritoneal cavities were washed with 10 ml of PBS. The lavage fluids were analysed for PMN recruitment by FACS using the neutrophil marker Ly-6G (1A8, BD Biosciences).

Mice and reagents. NALP3 targeting vector (Supplementary Fig. 1) was electroporated into C57BL/6 embryonic stem (ES) cells (Ozgene). Homologous recombinant ES cells were identified by Southern blot analysis and microinjected into C57BL/6 blastocysts. Offspring were backcrossed to C57BL/6 mice and germline transmission was confirmed by PCR of tail genomic DNA. Additional details on mice, preparation of crystals and reagents are given in the Supplementary Information.

Received 19 November; accepted 12 December 2005.

Published online 11 January 2006.

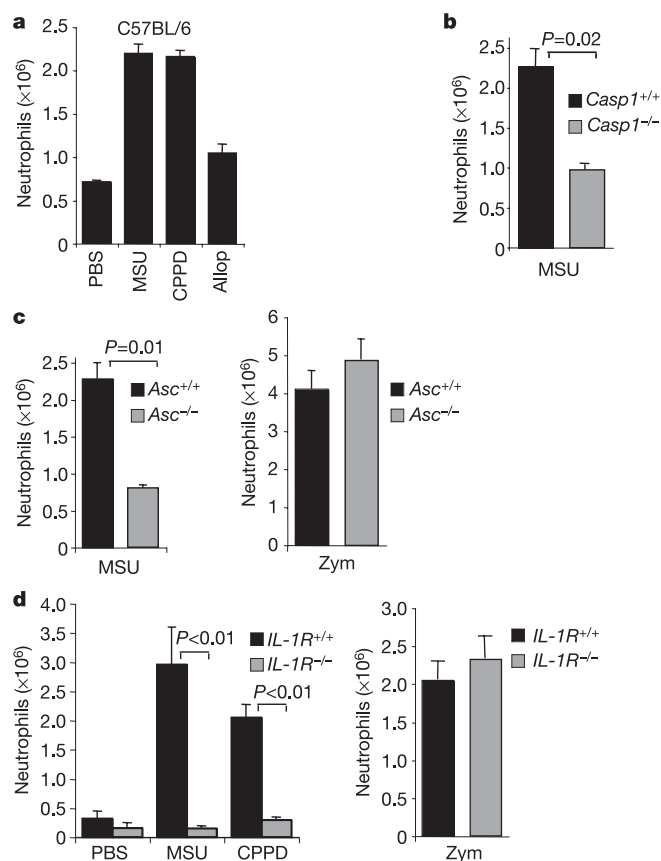


Figure 4 | Role of the inflammasome in a mouse model of crystal-mediated peritonitis. **a–d**, The indicated wild-type or mutant mice received 0.5 ml (intraperitoneally) of sterile PBS alone or supplemented with 1 mg of the indicated crystals or 0.2 mg of zymosan. Neutrophil influx was quantified 6 h later (values are $\pm$ s.e.m. of $n = 4$ –6 mice per group). Unpaired Student's *t*-test was used to calculate *P* values. Allop, allopurinol.

- Wollaston, H. W. On gouty and urinary concretions. *Phil. Trans.* **87**, 386–400 (1797).
- Shi, Y., Evans, J. E. & Rock, K. L. Molecular identification of a danger signal that alerts the immune system to dying cells. *Nature* **425**, 516–521 (2003).
- Galon, J., Aksentijevich, I., McDermott, M. F., O'Shea, J. J. & Kastner, D. L. TNFRSF1A mutations and autoinflammatory syndromes. *Curr. Opin. Immunol.* **12**, 479–486 (2000).
- Martinon, F. & Tschopp, J. Inflammatory caspases: linking an intracellular innate immune system to autoinflammatory diseases. *Cell* **117**, 561–574 (2004).
- Stojanov, S. & Kastner, D. L. Familial autoinflammatory diseases: genetics, pathogenesis and treatment. *Curr. Opin. Rheumatol.* **17**, 586–599 (2005).
- Dinarello, C. A. Blocking IL-1 in systemic inflammation. *J. Exp. Med.* **201**, 1355–1359 (2005).
- Shoham, N. G. et al. Pyrin binds the PSTPIP1/CD2BP1 protein, defining familial Mediterranean fever and PAPA syndrome as disorders in the same pathway. *Proc. Natl Acad. Sci. USA* **100**, 13501–13506 (2003).
- Chae, J. J. et al. Targeted disruption of pyrin, the FMF protein, causes heightened sensitivity to endotoxin and a defect in macrophage apoptosis. *Mol. Cell* **11**, 591–604 (2003).
- Burns, K., Martinon, F. & Tschopp, J. New insights into the mechanism of IL-1 β maturation. *Curr. Opin. Immunol.* **15**, 26–30 (2003).
- Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-1 β . *Mol. Cell* **10**, 417–426 (2002).
- Mariathasan, S. et al. Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
- Tschopp, J., Martinon, F. & Burns, K. NALPs: a novel protein family involved in inflammation. *Nature Rev. Mol. Cell Biol.* **4**, 95–104 (2003).
- Martinon, F., Agostini, L., Meylan, E. & Tschopp, J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. *Curr. Biol.* **14**, 1929–1934 (2004).
- Martinon, F. & Tschopp, J. NLRs join TLRs as innate sensors of pathogens. *Trends Immunol.* **26**, 447–454 (2005).
- Yamamoto, M. et al. ASC is essential for LPS-induced activation of procaspase-1 independently of TLR-associated signal adaptor molecules. *Genes Cells* **9**, 1055–1067 (2004).
- Faires, J. S. & McCarty, D. J. Acute arthritis in man and dog after intrasynovial injection of sodium urate crystals. *Lancet* **280**, 682–685 (1962).
- Dalbeth, N. & Haskard, D. O. Mechanisms of inflammation in gout. *Rheumatology (Oxford)* **44**, 1090–1096 (2005).
- Meng, Z. H., Hudson, A. P., Schumacher, H. R. Jr, Baker, J. F. & Baker, D. G. Monosodium urate, hydroxyapatite, and calcium pyrophosphate crystals induce tumour necrosis factor- α expression in a mononuclear cell line. *J. Rheumatol.* **24**, 2385–2388 (1997).

19. Chapman, P. T. *et al.* Endothelial activation in monosodium urate monohydrate crystal-induced inflammation: *in vitro* and *in vivo* studies on the roles of tumor necrosis factor alpha and interleukin-1. *Arthritis Rheum.* **40**, 955–965 (1997).
20. Hoffman, H. M. *et al.* Prevention of cold-associated acute inflammation in familial cold autoinflammatory syndrome by interleukin-1 receptor antagonist. *Lancet* **364**, 1779–1785 (2004).
21. Hawkins, P. N., Lachmann, H. J. & McDermott, M. F. Interleukin-1-receptor antagonist in the Muckle–Wells syndrome. *N. Engl. J. Med.* **348**, 2583–2584 (2003).
22. Molad, Y. Update on colchicine and its mechanism of action. *Curr. Rheumatol. Rep.* **4**, 252–256 (2002).
23. Malawista, S. E. & Seegmiller, J. E. The effect of pretreatment with colchicine on the inflammatory response to microcrystalline urate: A model for gouty inflammation. *Ann. Intern. Med.* **62**, 648–657 (1965).
24. Getting, S. J. *et al.* Molecular determinants of monosodium urate crystal-induced murine peritonitis: a role for endogenous mast cells and a distinct requirement for endothelial-derived selectins. *J. Pharmacol. Exp. Ther.* **283**, 123–130 (1997).
25. Goldfinger, S. E., Howell, R. R. & Seegmiller, J. E. Suppression of metabolic accompaniments of phagocytosis by colchicine. *Arthritis Rheum.* **8**, 1112–1122 (1965).
26. Liu-Bryan, R., Scott, P., Sydlaske, A., Rose, D. M. & Terkeltaub, R. Innate immunity conferred by Toll-like receptors 2 and 4 and myeloid differentiation factor 88 expression is pivotal to monosodium urate monohydrate crystal-induced inflammation. *Arthritis Rheum.* **52**, 2936–2946 (2005).
27. Janeway, C. A. Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
28. Matzinger, P. The danger model: a renewed sense of self. *Science* **296**, 301–305 (2002).
29. Hoffman, H. M., Mueller, J. L., Broide, D. H., Wanderer, A. A. & Kolodner, R. D. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle–Wells syndrome. *Nature Genet.* **29**, 301–305 (2001).
30. Agostini, L. *et al.* NALP3 forms an IL-1 β -processing inflammasome with increased activity in Muckle–Wells autoinflammatory disorder. *Immunity* **20**, 319–325 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Mattmann for technical support and A. So, H. Everett, E. Meylan, M. Thome and P. Schneider for discussions and critical reading of the manuscript. We thank S. Mariathasan, V. M. Dixit, R. A. Flavell, M. Kopf and S. Akira for the gift of various knockout mice. This work was supported by grants from the Swiss National Science Foundation and the Commission of Technology and Innovation (CTI). V.P. is supported by a fellowship of the FRM (Fondation pour la Recherche Médicale); A.T. by a NCCR grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.T. (jurg.tschoep@unil.ch).

LETTERS

Expression profiling in primates reveals a rapid evolution of human transcription factors

Yoav Gilad^{1†}, Alicia Oshlack², Gordon K. Smyth², Terence P. Speed^{2,3} & Kevin P. White¹

Although it has been hypothesized for thirty years that many human adaptations are likely to be due to changes in gene regulation¹, almost nothing is known about the modes of natural selection acting on regulation in primates. Here we identify a set of genes for which expression is evolving under natural selection. We use a new multi-species complementary DNA array to compare steady-state messenger RNA levels in liver tissues within and between humans, chimpanzees, orangutans and rhesus macaques. Using estimates from a linear mixed model, we identify a set of genes for which expression levels have remained constant across the entire phylogeny (~70 million years), and are therefore likely to be under stabilizing selection. Among the top candidates are five genes with expression levels that have previously been shown to be altered in liver carcinoma. We also find a number of genes with similar expression levels among non-human primates but significantly elevated or reduced expression in the human lineage, features that point to the action of directional selection. Among the gene set with a human-specific increase in expression, there is an excess of transcription factors; the same is not true for genes with increased expression in chimpanzee.

A number of recent studies have used DNA microarrays to compare patterns of gene expression between closely related species^{2–9}. Within primates, the focus has been primarily on human–chimpanzee comparisons, estimating gene expression profiles for a number of tissues, including liver, brain and heart^{2,6,7,10}. The aim has been to characterize general trends in the evolution of gene expression rather than to identify specific genes of interest. To date, conclusions about the selection pressures acting on gene expression have been conflicting^{2,3,6,11–13}.

These studies have all relied on data collected from arrays using gene probes that were designed on the basis of human sequences only. However, sequence mismatches affect hybridization intensity and can therefore bias estimates of gene expression differences between species¹⁴. This limitation of single-species arrays is especially problematic when the goal is to study how expression changes over evolutionary time. To make comparisons between more distantly related primate species, we generated a multi-species cDNA array that allows comparison of gene expression between species without the confounding effects of sequence divergence¹⁴. This cDNA array contains probes for 1,056 orthologous genes from four species (see Supplementary Methods)¹⁴.

We used this array to compare gene expression profiles in the livers of humans, chimpanzees (*Pan troglodytes*), orangutans (*Pongo pygmaeus*) and rhesus macaques (*Macaca mulatta*), the phylogeny of which represents approximately 70 million years (Myr) of evolution. By assigning expression changes in the liver to particular lineages, we were able to identify the first set of genes for which regulation seems to be under lineage-specific selection pressures. In order to measure

gene expression levels within and between species, we extracted RNA from liver samples of five adult males from each of the four species. A common reference design was used, with a sixth human liver sample serving as the reference. We performed four technical replicates of each comparison, for a total of 80 hybridizations. Results from all species were obtained for 907 genes, used in subsequent analyses (Supplementary Table S1).

After image analysis, background correction and normalization, the log expression values were analysed using a linear mixed model with fixed effects for species and sequence mismatches, and a random effect for individuals within species (see Methods). For each gene, we used residual maximum likelihood¹⁵ to estimate the fixed effects and variances. Hypothesis testing was performed using likelihood ratio tests (see Methods).

As a first step, we identified genes that are differentially expressed between species (Table 1). A phylogenetic tree based on the number of differentially expressed genes between species¹⁶ recapitulates their known phylogeny (Supplementary Fig. S1). However, the number of significantly differentially expressed genes does not always increase with evolutionary time.

Focusing on human and chimpanzee, we found 110 genes (12%) to be differentially expressed at a false discovery rate (FDR)¹⁷ of 0.01, with a mean absolute log ratio of 1.56-fold difference (Supplementary Table S2). Our observation is in general agreement with a statistical meta-analysis¹¹ of the data from ref. 2. In contrast to this meta-analysis, however, we find that equal numbers of genes have elevated (55) or reduced (55) expression levels in humans compared to chimpanzees.

To estimate lineage-specific changes in expression levels, we used the expression profiles from orangutan and rhesus macaques as outgroups for 84 of the genes that show significantly different expression between human and chimpanzee (Fig. 1a; see Methods). Using this approach, we found similar numbers of genes for which expression has been altered in either the human or the chimpanzee lineage. Moreover, in both species, the numbers of genes that show increased or decreased expression levels relative to the estimated ancestral expression level is similar (45 and 43 of the genes are upregulated in humans and chimpanzees, respectively). In addition, the average or median fold change in gene expression level is similar regardless of the lineage or the trend (that is, up or down)

Table 1 | Inter-species differentially expressed genes

	Chimpanzee	Orangutan	Rhesus macaque
Human	110	128	176
Chimpanzee	–	150	141
Orangutan	–	–	129

¹Department of Genetics and Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06510, USA. ²Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. ³Department of Statistics and Program in Biostatistics, University of California, Berkeley, California 94720, USA. †Present address: Department of Human Genetics, University of Chicago, Chicago, Illinois 60605, USA.

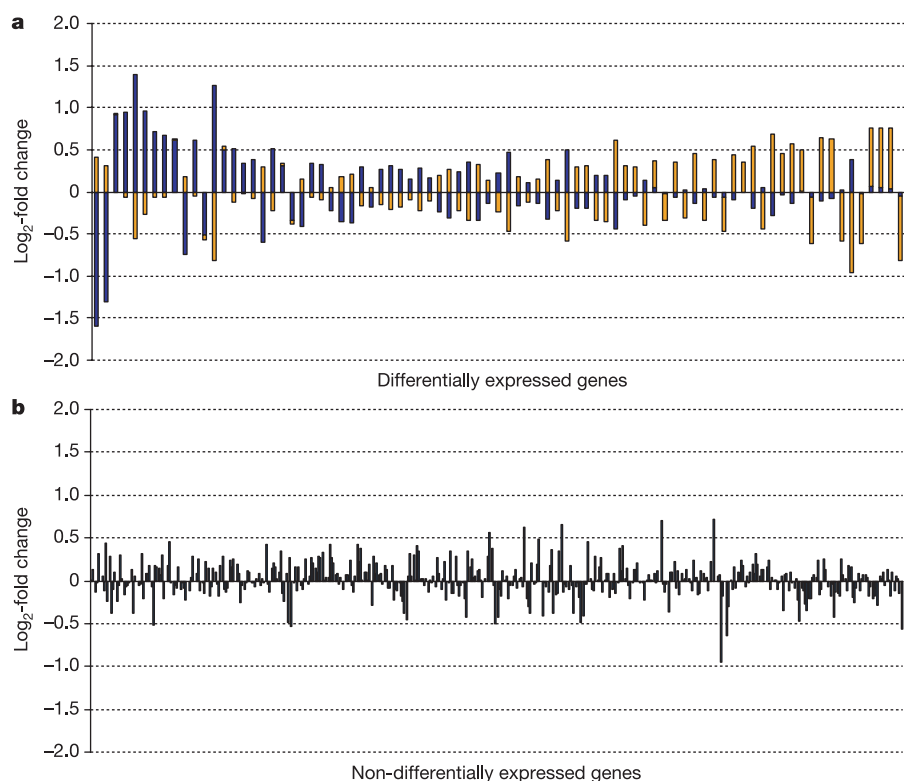


Figure 1 | Expression changes in specific lineages. **a**, For 84 genes that are differentially expressed between human and chimpanzee, the log₂-fold change relative to the common ancestor is given for the human (blue) and chimpanzee (orange) lineages. Genes are ordered by the ratio of their expression changes in the human lineage compared to the chimpanzee

lineage. **b**, For 446 genes that are not differentially expressed between human and chimpanzee, and for which an ancestral state could be estimated (see Methods), the log₂-fold change relative to the common ancestor is shown for the human lineage.

(Supplementary Fig. S2). The pattern also holds for expression changes in the human lineage in genes that are not differentially expressed between human and chimpanzee (Fig. 1b; 52% of the genes were upregulated). These observations do not agree with previous studies^{2,10}. Possible explanations for the discrepancy are the use of human microarrays for inter-primate comparisons², or the assignment of expression changes to lineages in the absence of outgroup data¹⁰.

Our approach also allows the identification of genes for which regulation is likely to have evolved under stabilizing selection. Previous studies have done this by testing for deviations from neutrality or stabilizing selection^{13,16,18}. Such an approach requires a model for the evolution of expression, and thus relies on a number of parameter estimates about which there is considerable uncertainty in primates (for example, the neutral expression change per generation, and the environmental and mutational variance for each gene). Instead of specifying an explicit model, we used statistical analyses to rank genes according to their pattern of evolutionary change among the four species, and focused on those at the top of the list as the most promising candidates.

First, we identified genes that best fitted a model of constant expression level throughout the phylogeny, reasoning that these represent promising candidates for stabilizing selection. A majority of the genes on the array (60%) do not show significant inter-species expression differences. However, failure to reject the null hypothesis of no expression difference between species can result from constant expression level in all individuals in all species (Fig. 2a) or large within-species variance (Fig. 2b)—especially as primate tissues cannot be staged¹⁰. As our aim is to identify genes under stabilizing selection, we are only interested in the former scenario. We therefore ranked genes by their expression variation among individuals across all species (see Methods). Genes at the top of our list are not

significantly differentially expressed between species, and also have low within-species variance (Fig. 2). The expression levels of these genes seem to have remained constant for ~70 Myr¹⁹, suggesting that their regulation is under evolutionary constraint. Among the first 100 genes on our list (Supplementary Table S3), the most significant enrichment ($P < 10^{-8}$; uncorrected for multiple tests) is for genes from the category 'regulation of cellular physiological process' (Gene Ontology ID 0051244; <http://www.geneontology.org>). As we expect transcription of such genes to be similar across individuals and species, this finding serves as a validation of the approach.

A number of recent papers have argued that the majority of expression differences observed between primates are neutral, based primarily on the observation that the mean square fold change in expression levels in liver and brain increases linearly with species divergence time^{6,12}. Having found no clear increase in the number of significantly differentially expressed genes with time (Table 1), we re-examined the mean square fold change for our data. This revealed no linear increase over time (Supplementary Fig. S3). Moreover, our observation that many genes show stable expression levels over 70 Myr suggests that, rather than evolving mostly neutrally, expression levels are often under stabilizing selection, consistent with findings in *Drosophila*^{16,18} and in *C. elegans*²⁰.

This finding has implications for studies of human disease. Indeed, our observations suggest that many changes in gene regulation may be deleterious and hence influence disease susceptibility. Consistent with this, among the top 100 genes for which regulation is probably evolving under stabilizing selection, genes associated with human cancer are slightly enriched (9% compared to 5% in the total gene sample; $P = 0.10$, one-tailed Fisher's exact test). Moreover, the expression levels of five genes (*MBD4*, *WWOX*, *ING1*, *ATP7B* and *IGFBP2*; ranked 5th, 12th, 28th, 58th and 66th, respectively) have been shown to be altered specifically in liver carcinoma^{21–24}. These

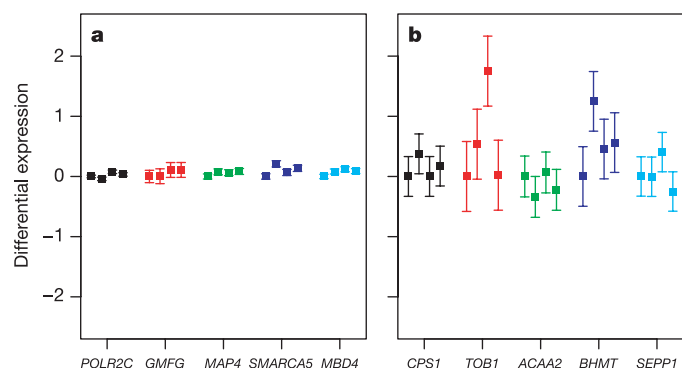


Figure 2 | Genes that are not differentially expressed across species. In each plot, different genes (x-axis) are represented by different colours. For each gene, the estimated expression level ($\pm$ s.e.m.) is shown for humans, chimpanzees, orangutans and rhesus macaque (left to right). **a**, The five highest-ranked genes (see Methods). These genes have constant expression levels in all species, suggesting that their expression levels are under stabilizing selection. **b**, Examples of genes that are not differentially expressed across species, probably due to high within-species variance (gene rankings 489–493).

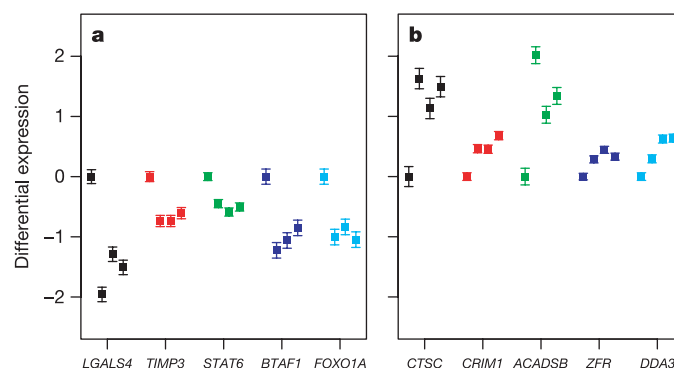


Figure 3 | Genes with distinct expression pattern in humans. Different genes (x-axis) are represented by distinct colours. For each gene, the $\log_2$ expression levels for humans are set to zero. Estimated gene expression level relative to human ($\pm$ s.e.m.) is shown for humans, chimpanzees, orangutans and rhesus macaque (left to right). Shown are examples of five genes that are not differentially expressed in the non-human primates but are upregulated (**a**) or downregulated (**b**) in humans. The expression levels of these genes seem to have been under stabilizing selection in the non-human primates and under directional selection in the human lineage.

findings suggest that focusing on genes with conserved expression levels among primates may be helpful in identifying promising candidates for disease-association studies, much like phylogenetic shadowing of DNA sequences²⁵ can aid in the identification of non-coding elements of functional importance.

Using this general approach, we also identified genes for which expression levels are not significantly different among non-human primates but are significantly elevated or reduced in humans relative to each of the three other species (see Methods and Supplementary Table S4). In other words, the expression level of the gene has remained similar over ~65 Myr of evolution and then changed over the ~5 Myr of the human lineage, indicative of directional selection in humans. Our analysis revealed 14 genes with significantly higher expression levels in humans and five with lower expression (Fig. 3). We note that we are likely to be missing a number of targets of positive selection: gene expression varies across tissues and developmental stages²⁶, and as a result, the absence of support for selection in primate expression data is weak evidence against it.

Notably, among the genes with higher expression in humans, we find a significant excess of transcription factors (5/12, 42% compared with 10% representation on the array; $P = 0.003$ by Fisher's exact test, including all genes for which GO annotation was available), whereas no transcription factors were found among genes with unusually low expression in humans. We repeated this analysis using a less stringent criterion to identify genes for which the mean expression level in humans differed significantly from that of non-human primates (see Methods). Again, transcription factors were overrepresented among the 30 genes with elevated expression in humans (30%; $P = 0.001$, Fisher's exact test), and no transcription factors were found among 19 genes with reduced expression. In contrast, when these analyses were applied to chimpanzee (Supplementary Table S5), the number of transcription factors was equivalent among genes with elevated (9%) or reduced (9%) expression levels (for the less stringent cutoff), and neither proportion was significantly different from the overall representation on the array (that is, 10%). It is unlikely that these observations can be explained by differential degradation of transcripts encoding specific classes of proteins²⁷, as no difference in RNA quality was observed between human and non-human primate samples during sample preparation (on the basis of electrophoretic analyses).

In addition to the rapid evolution of expression levels, genes encoding transcription factors have also been shown to evolve rapidly in the human lineage at the coding sequence level²⁸. Together, these

findings raise the possibility that the function and regulation of transcription factors have been substantially modified in the human lineage, potentially affecting many downstream targets over a short evolutionary time frame. Notably, the opposite finding emerged from studies of closely related *Drosophila* species, in which the expression levels of transcription factors were shown to evolve slower than genes encoding other types of proteins^{16,18}. Given the large number of phenotypic changes in the human lineage¹, it is tempting to speculate that relative rates of transcription factor evolution may serve as an indicator of rates of phenotypic evolution at the organismal level.

Finally, to examine the extent to which evolution of protein-coding regions mirrors gene expression level changes in the liver, we considered three sets of genes: those for which expression levels seem to be under directional selection in humans (set A), the top 100 candidates for stabilizing selection (set B) and the remaining genes (set C). To assess the evidence for natural selection acting on coding regions, we used estimates of the posterior probability that a gene is subject to positive or negative selection based on synonymous and non-synonymous nucleotide polymorphism and divergence levels at genes on our array²⁸. Using this approach (with a posterior probability of 0.05), only 6% of genes in set C and 4% in set B are inferred to evolve under positive selection. In contrast, among set A, significantly more genes (25%) are inferred to evolve under positive selection ($P = 0.03$, one-tailed Fisher's exact test). These observations suggest that genes with expression levels under directional selection in humans are somewhat more likely to show accelerated amino acid evolution.

In summary, the use of a new multi-species cDNA array has allowed us to identify a set of genes with regulation under natural selection in humans. In particular, the over-representation of transcription factors among the genes with modified expression levels in the human lineage is consistent with the suggestion that most differences between human and chimpanzee are due to changes in gene regulation¹, and might provide insight into their genetic architecture.

METHODS

Study design and analysis. The 80 arrays were scanned using a GenePix Axon scanner and data were extracted using GenePix 6 (Molecular Devices) to give Cy5 and Cy3 foreground and background fluorescence intensities. Analysis was done in the R computing environment (<http://www.r-project.org>). Background-corrected Cy5 and Cy3 intensities were produced using the 'normexp' method

with an offset of 50, implemented in the limma software package²⁹. Lowess curves for intensity-dependent normalization were generated in a way similar to ref. 14, where probes from the two species involved in the hybridization were used to fit the curves. All probes on the array were adjusted by the fitted lowess curve (see Supplementary Methods). We concentrated on the 907 genes on the array for which successful polymerase chain reaction (PCR) products were obtained from all species¹⁴. The expression log ratios for each gene were analysed using the linear mixed model:

$$y_{tijp} = \mu_t + \kappa_{tp} - \kappa_{hp} + \alpha_{ti} + \varepsilon_{tijp}$$

in which we have suppressed the gene labels. Here, y_{tijp} is the normalized log₂ ratio measured for target species t for replicate j of individual i on species probe p . The term μ_t is the expected log ratio of the expression level of the gene in target species t relative to the human reference, and κ_{tp} and κ_{hp} are parameters corresponding to the reduction in the log expression levels caused by reduced affinity owing to target and probe sequence mismatches. As each hybridization has target species t on the red channel and the human reference on the green channel, there are two κ terms for each measurement. We assume that κ_{ti} is equal to 0, and that the affinity adjustments are symmetrical in target and probe (that is, $\kappa_{tp} = \kappa_{pt}$). The term α_{ti} is the random effect for individual i of species t , assumed to be uncorrelated with mean zero and variance σ_α^2 . Finally, ε_{tijp} is the residual error term, and these are assumed to be uncorrelated with mean zero and variance σ_ε^2 . We also considered models that included random effects for probes within arrays and a crossed term for an array $\times$ probe interaction, but found that the contributions from these terms were substantially smaller than the error term and therefore did not warrant inclusion in the model. (See Supplementary Information for further details on the parameters and model.) For each gene, the model was fitted by residual maximum likelihood using statmod and lme software packages³⁰.

Hypothesis testing. Likelihood ratio tests were used for hypothesis testing. Under the full model, for each gene, 12 parameters (4 μ_t parameters, 6 κ_{tp} parameters, σ_α^2 and σ_ε^2) were estimated by maximum likelihood. Genes deemed to be under stabilizing selection were those for which the fit of a reduced model with $\mu = \mu_h = \mu_c = \mu_o = \mu_r$ was adequate (h , human; c , chimpanzee; o , orangutan; r , rhesus macaque). Such genes were selected on the basis of the likelihood ratio test statistic comparing the fit under this sub-model to that under the full model. Under the null hypothesis, $-2(\log\text{-likelihood ratio})$ has an approximate χ^2 distribution on 3 degrees of freedom, and genes for which this statistic was less than 12.4 ($P = 6.1 \times 10^{-3}$) were chosen. We then ranked these genes according to the magnitude of the between-to-within individual ratio mean squares ($16\hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2 / \hat{\sigma}_\varepsilon^2$) starting with genes for which this was small. We note that the latter process alone would not suffice to identify genes that are not differentially expressed between species (Supplementary Fig. S4).

To select genes that were different in human compared to the other three species, we combined three criteria. First, we used a likelihood ratio statistic to exclude genes that were differentially expressed in the non-human primates. We maximized the likelihood under the constraints $\mu = \mu_c = \mu_o = \mu_r$, constructed the ratio of this likelihood compared to the full model, and removed genes where we estimated significant differences. Second, we used a likelihood ratio statistic to rank genes on the basis of differences between human and the other species (that is, $\mu \neq \mu_h$). We chose a cutoff statistic of 16 ($P = 6.3 \times 10^{-5}$) to select genes, but also investigated genes selected under a more relaxed cutoff of 12, which corresponds to $\sim 1\%$ FDR¹⁷. Third, we restricted the list to genes with small between relative to within individual variance. Pairwise differences between species were also constructed using a likelihood ratio statistic with a cutoff chosen to give 1% FDR, assuming a χ^2_1 distribution (numbers are given in Table 1). We found by simulations that the null likelihood ratio test statistic was well approximated by a χ^2 distribution, implying that our assumptions are accurate (data not shown).

We note that correlation between species due to shared phylogeny is not expected to influence our results, as no structure is imposed on the parameters for the means of the different species and no model is fitted to them across species.

Received 5 November; accepted 29 December 2005.

- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
- Caceres, M. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
- Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745 (2003).
- Karaman, M. W. *et al.* Comparative analysis of gene-expression patterns in

human and African great ape cultured fibroblasts. *Genome Res.* **13**, 1619–1630 (2003).

- Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, E132 (2004).
- Khaitovich, P. *et al.* Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462–1473 (2004).
- Rise, M. L. *et al.* Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res.* **14**, 478–490 (2004).
- Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. & McIntyre, L. M. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* **21**, 1308–1317 (2004).
- Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854 (2005).
- Hsieh, W. P., Chu, T. M., Wolfinger, R. D. & Gibson, G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**, 747–757 (2003).
- Khaitovich, P., Paabo, S. & Weiss, G. Toward a neutral evolutionary model of gene expression. *Genetics* **170**, 929–939 (2005).
- Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution Int. J. Org. Evolution* **59**, 126–137 (2005).
- Gilad, Y., Rifkin, S. A., Bertone, P., Gerstein, M. & White, K. P. Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res.* **15**, 674–680 (2005).
- Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).
- Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
- Rifkin, S. A., Houle, D., Kim, J. & White, K. P. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
- Glazko, G. V. & Nei, M. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003).
- Denver, D. R. *et al.* The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genet.* **37**, 544–548 (2005).
- Park, S. W. *et al.* Frequent downregulation and loss of WWOX gene expression in human hepatocellular carcinoma. *Br. J. Cancer* **91**, 753–759 (2004).
- Sugeno, H. *et al.* Expression of copper-transporting P-type adenosine triphosphatase (ATP7B) in human hepatocellular carcinoma. *Anticancer Res.* **24**, 1045–1048 (2004).
- Zhu, Z. *et al.* Inhibitory effect of tumour suppressor p33^{ING1b} and its synergy with p53 gene in hepatocellular carcinoma. *World J. Gastroenterol.* **11**, 1903–1909 (2005).
- Chiba, T. *et al.* Identification and investigation of methylated genes in hepatoma. *Eur. J. Cancer* **41**, 1185–1194 (2005).
- Boffelli, D. *et al.* Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391–1394 (2003).
- Ludwig, M. Z. *et al.* Functional evolution of a cis-regulatory module. *PLoS Biol.* **3**, e93 (2005).
- Yang, E. *et al.* Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res.* **13**, 1863–1872 (2003).
- Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).
- Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. & Huber, W.) 397–420 (Springer, New York, 2005).
- Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer-Verlag, New York, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. E. Holt for pointing out a possible technical explanation for the excess of upregulated transcription factors in humans, the Yale hospitals, S. Paabo and the Yerkes Primate Center for providing samples used in the study, and A. Clark and M. Przeworski for comments on the manuscript. This research was supported by grants to K.P.W. from the W. M. Keck Foundation, the Arnold and Mabel Beckman Foundation and the National Human Genome Research Institute of the National Institutes of Health, and by NHMRC grants to G.K.S., T.P.S. and A.O. Y.G. was supported by an EMBO fellowship.

Author Information Expression data from this study have been deposited in the GEO database under the series accession number GSE2569. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.G. (gilad@uchicago.edu) or K.P.W. (kevin.white@yale.edu).

LETTERS

Nanospring behaviour of ankyrin repeats

Gwangrog Lee¹, Khadar Abdi², Yong Jiang¹, Peter Michaely³, Vann Bennett² & Piotr E. Marszalek¹

Ankyrin repeats are an amino-acid motif believed to function in protein recognition; they are present in tandem copies in diverse proteins in nearly all phyla¹. Ankyrin repeats contain antiparallel α -helices that can stack to form a superhelical spiral². Visual inspection of the extrapolated structure of 24 ankyrin-R repeats² indicates the possibility of spring-like behaviour of the putative superhelix. Moreover, stacks of 17–29 ankyrin repeats in the cytoplasmic domains of transient receptor potential (TRP) channels have been identified as candidates for a spring that gates mechanoreceptors in hair cells as well as in *Drosophila* bristles^{3–5}. Here we report that tandem ankyrin repeats exhibit tertiary-structure-based elasticity and behave as a linear and fully reversible spring in single-molecule measurements by atomic force microscopy. We also observe an unexpected ability of unfolded repeats to generate force during refolding, and report the first direct measurement of the refolding force of a protein domain. Thus, we show that one of the most common amino-acid motifs has spring properties that could be important in mechanotransduction and in the design of nanodevices.

The atomic structure of 12 ankyrin-R repeats suggests that ankyrin stacks composed of $n \geq 24$ repeats should form a full superhelical

turn with putative spring properties^{2,3}. We used an atomic force microscope (AFM) to identify individual stacks of 24 ankyrin-B repeats (Supplementary Fig. S2) and found that they do indeed have a hook-like shape² with the molecules' end-to-end distance closely matching the ~ 12 nm determined for the extrapolated structure² (Fig. 1a). Thus, the AFM images strongly suggest that the engineered protein, bearing at its terminus a glutathione S-transferase (GST) module, is correctly folded and does not aggregate. These conclusions are further supported by circular dichroism and hydrodynamic measurements (Supplementary Table 1 and Supplementary Fig. S1).

For elasticity measurements, heptahistidine-tagged polypeptides containing 24 ankyrin-B repeats with or without GST, or 12 repeats with GST, were immobilized on a glass surface bearing the metal chelate *N*-nitrilotriacetic acid (NTA)^{6,7} (Fig. 1a). Molecules were stretched vertically, in solution, by the AFM cantilever, and their length and tension were measured with subnanometre and ~ 10 pN precision^{8–10}. Most trials revealed complex force–extension profiles with irregularly spaced force peaks typical of multiple molecules (Supplementary Fig. S4a). However, $\sim 5\%$ of the force–extension curves had simple and consistent features that, we argue, represent

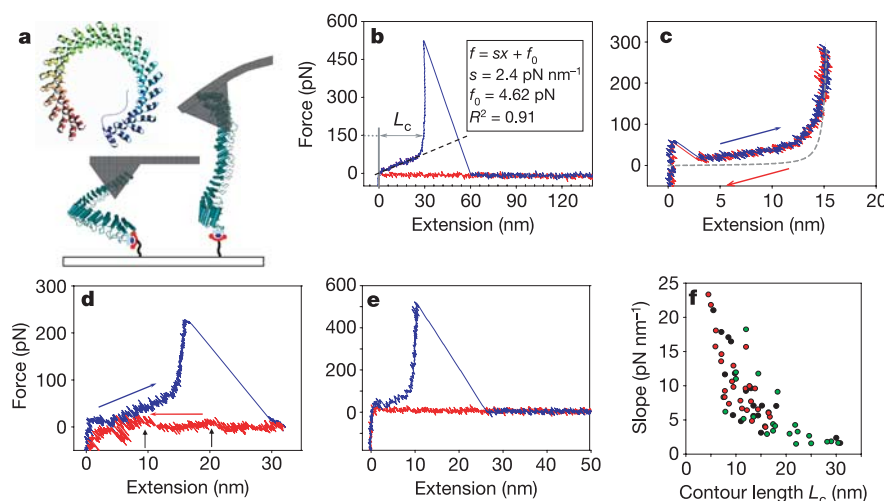


Figure 1 | Atomic force microscopy measurements reveal the linear elasticity of ankyrin-B repeats. **a**, The extrapolated structure of 24 ankyrin-R repeats² and a diagram of the elasticity measurement on a His-tagged ankyrin fragment bound to NTA (red handles) and stretched with the AFM cantilever. **b–e**, Force–extension curves of individual ankyrins: 24 repeats with GST (**b–d**); 24 repeats with no GST (**e**). In **b**, the molecule detached from the AFM at the stretching force peak (blue); in **c**, the molecule's attachment was preserved after stretching (blue trace) and the molecule was subsequently relaxed (red trace); in **d**, the molecule unfolded at the stretching force peak (blue) but remained attached to the AFM. The inset in

b shows how ankyrin spring constant s was determined by fitting the straight-line equation to the data: $f = sx + f_0$, force; x , extension; R^2 , coefficient of determination; L_c , contour length (determined as the extension at 150 pN). In **d**, the relaxing trace (red) after the unfolding of ankyrin that occurred at the stretching force peak (blue) revealed small force peaks (arrows) indicating the refolding of individual repeats (see also Supplementary Fig. S6). **f**, Ankyrin spring constant decreases with the increasing length of the stretched fragment. Black dots, 24 repeats with GST; green dots, 24 repeats with no GST; red dots, 12 repeats with GST.

¹Department of Mechanical Engineering and Materials Science and Center for Biologically Inspired Materials and Material Systems, Duke University, Durham, North Carolina 27708, USA. ²Howard Hughes Medical Institute and Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27708, USA. ³Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

the elasticity of single ankyrin stacks. Figure 1b–e shows 4 of 77 similar recordings. Interestingly, at forces of 100 pN or less, all of these curves display a linear region where force is directly proportional to extension. The length L_c of the stretched segment (Fig. 1b, inset) varied between ~ 5 and ~ 30 nm, which is consistent with random adsorption of the AFM tip along the ankyrin polypeptide. Provided that stretching was stopped before unfolding, the extension was fully reversible (compare the red and blue traces in Fig. 1c). In other cases where the molecules detached (Fig. 1b, e) or were unfolded (Fig. 1d), extension was not reversible.

The fact that Fig. 1b, d, e displays a single force peak strongly indicates that these recordings were obtained on single molecules. Note also that the relaxing trace in Fig. 1d (red curve) shows small refolding force peaks (arrows) generated by individual ankyrin repeats (see below and Supplementary Fig. S6). Moreover, the force curves shown in Fig. 2b, c show an initial linear phase followed by rupture of the repeat stack and unfolding of individual repeats (see below). Together, these traces provide direct evidence that the unusual linear elasticity captured in our AFM recordings on ankyrin do indeed represent the properties of individual molecules.

Unlike most modular proteins studied by atomic force microscopy, including titin^{8,10–13}, tenascin⁹, fibronectin¹⁴, ubiquitin¹⁵, spectrin¹⁶ and filamin¹⁷, whose elasticity typically follows a highly nonlinear entropic behaviour exemplified by the worm-like chain (WLC) model¹⁸ (Fig. 1c, dashed line), ankyrin is a hookean or linear spring for which the tension is proportional to the extension. To behave similarly to ankyrin, a WLC polymer would need an unphysical persistence length of less than 1 Å.

Elastic properties of ankyrin repeats have been proposed on the basis of their coil-like shape^{3,19,20}. Recent steered molecular dynamics (SMD) simulations also predict that ankyrin-R repeats stretch reversibly by straightening the superhelical turn, when subjected to forces of ~ 25 –100 pN. These simulations predicted that the spring constant of ankyrin-R is 4.1 – 4.7 pN nm⁻¹ for 24 repeats and increases to 16.4 pN nm⁻¹ for 12 repeats⁵, indicating that the stiffness of the stack increases with a decreasing number of repeats. We plotted the spring constants of various ankyrin constructs as a function of L_c (Fig. 1f) and found that longer ankyrin stacks are on average more compliant than shorter ones, which is consistent with expectations of a hookean spring and the SMD simulations⁵. Moreover, the GST-tagged protein does not contribute to the spring constant. The range of spring constants is 1.5 – 21 pN nm⁻¹ for 24 ankyrin-B repeats and 4.0 – 23 pN nm⁻¹ for the 12-repeat construct. The spring constant measured for the longest fragments (extensions: 25–30 nm; Fig. 1f) presumably containing all 24 ankyrin repeats, is 1.87 ± 0.31 pN nm⁻¹ (mean $\pm$ s.d., $n = 5$). For comparison, the estimate based on the SMD calculation gives 4.4 pN nm⁻¹ (ref. 5) and the spring constant of a spring that gates mechanoreceptors in hair cells was predicted to be 0.5 pN nm⁻¹ (ref. 21).

Ankyrin-B repeats extend reversibly in multiple stretch–relaxation cycles with no signs of hysteresis or wear over a large range of applied forces exceeding 100 pN (Fig. 1c). However, these stacks do unfold at relatively high forces ranging from 200 to 1,250 pN (365 ± 202 pN; mean $\pm$ s.d., $n = 32$; see, for example, Figs 1d and 2; Supplementary Fig. S9e). Figure 2 describes the mechanical unfolding of ankyrin stacks in detail. The first force peak, labelled A in Fig. 2a, is a rupture of a nonspecific adhesive bond between the AFM tip and the substrate. The next force peak of 310 pN reports the breakdown of the molecule and not its detachment from the tip (stage 1, peak labelled U) as demonstrated by the rest of the force curve, which captured the WLC-like elasticity (stage 2)²². From the length of the polypeptide chain liberated after the breakdown (U) we estimate that six repeats were unfolded simultaneously and subsequently stretched. Interestingly, further extension revealed a sawtooth pattern of six regularly spaced force peaks, each separated by 12.4 nm (on the basis of the WLC fits). This distance matches the length of an ankyrin repeat (33 residues $\times$ 0.36 nm per residue), indicating that

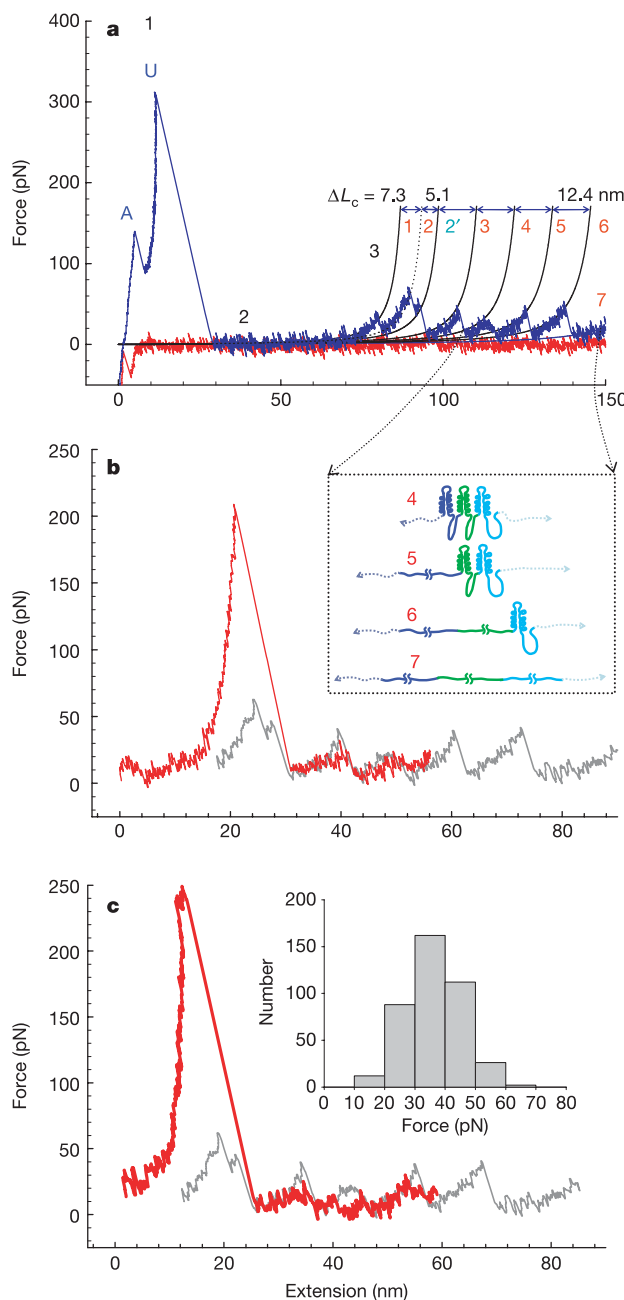


Figure 2 | Atomic force microscopy captures the mechanical breakdown of the ankyrin stack and the unfolding of individual repeats. **a**, The first force peak in the force–extension curve at ~ 150 pN (A) represents a rupture of a nonspecific adhesive bond. At 310 pN the stack breaks down (1, U); and about six repeats unravel cooperatively and are stretched in a WLC manner (curve labelled 2). At point 3 there is a sequential unfolding of the remaining repeats. Thin solid lines are the WLC fits to the data with a persistence length p of 2 nm and the contour length increment ΔL_c of 12.4 nm. Note that the second peak is actually composed of two peaks (2 and 2') separated from the first peak by 7.3 and 12.4 nm, respectively. **b**, A force–extension curve of another ankyrin fragment reveals in a single pull its linear elasticity, the breakdown of the stack at ~ 240 pN, and the unfolding of individual repeats (red trace). Inset, our interpretation of the events occurring in stage 3 of **a**. **c**, As in **b**, but obtained on a different ankyrin fragment. The inset shows a histogram of unfolding forces of individual repeats. The grey trace in **b** and **c** comes from the force–extension curve shown in **a**.

the force peaks might correspond to the sequential unfolding of individual ankyrin repeats. The force that is necessary to unfold a single ankyrin repeat (Figs 2 and 3; Supplementary Figs S9 and S10) is 37 ± 9 pN (mean $\pm$ s.d., $n = 404$ force peaks; Fig. 2c, inset), which

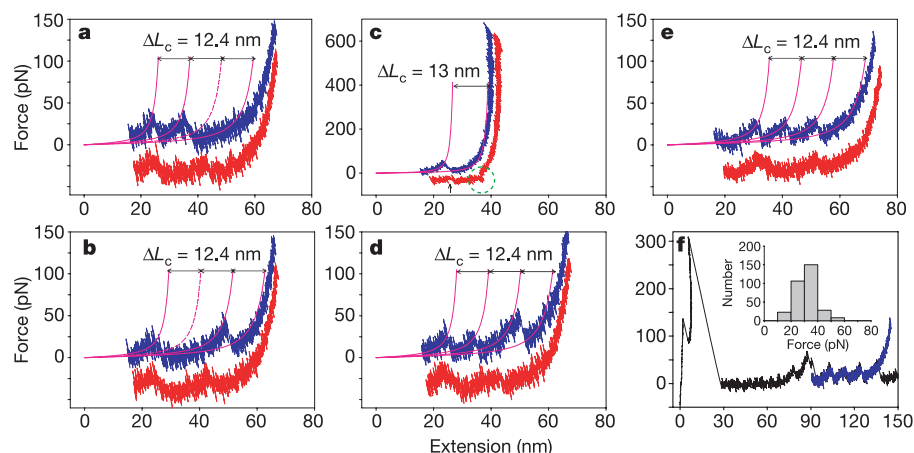


Figure 3 | Atomic force microscopy directly measures the refolding forces of ankyrin repeats. **a–e**, A sequence of stretch (blue traces) and release (red traces) measurements of a single partly unfolded ankyrin fragment (compare with Supplementary Fig. S7). Thin lines are the WLC fits with $p = 1$ nm and $\Delta L_c = 12.4$ nm. The relaxing traces are offset vertically for clarity. **a**, Cycle 4. **b**, Cycle 5. **c**, Cycle 6: two of the three previously unfolded repeats reattached to the stack and only a single repeat unfolded. Green

dotted circle, instability of the refolded stack captured in the relaxing trace. **d**, Cycle 7: the stretching trace reveals that the two repeats that had reattached to the stack now detached again and all three repeats unfolded sequentially, then refolded. **e**, As in **d**, but obtained after 26 more stretch–release cycles. **f**, The unfolding trace of **e** superimposes perfectly on the template trace of Fig. 2a. The inset shows a histogram of the refolding forces of individual repeats.

is similar to the unfolding force (25–35 pN) of individual spectrin domains¹⁶.

Figure 2b, c shows two additional examples of the forced unfolding of different ankyrin stacks, each captured in a single measurement. They clearly reveal the linear elasticity of the stack followed by its breakdown (high force peak), which in turn is followed by an unravelling of individual repeats (small force peaks). Significantly, these small force peaks overlap well with the small force peaks captured in Fig. 2a (shown in Fig. 2b, c as a grey trace), strongly indicating that all these force peaks represent the unfolding of individual repeats (see Fig. 2b, inset).

A remarkable feature of ankyrin repeats is their ability to refold in a process that reveals significant refolding forces (Fig. 1d). We further investigated refolding in a sequence of 33 consecutive stretch–release recordings obtained on a single ankyrin fragment (Fig. 3; see also Supplementary Figs S7–S9). The recordings in Fig. 3a–e were obtained on ankyrin that had already been partly unfolded and prestretched by 17 nm (Supplementary Fig. S7). Figure 3a (blue trace) revealed two small force peaks of 32 pN, separated by ~ 12 nm, similar to those in Fig. 2a, followed by the force curve, which was separated from the second peak by ~ 25 nm. This result indicates that the unfolding of a single ankyrin repeat might have been followed by the simultaneous unfolding of two repeats. The relaxing trace in Fig. 3a (red curve) also shows two force peaks of 27 pN and their spacing suggests that they report the refolding of a single repeat that was followed by the refolding of the two remaining repeats. The next cycle (Fig. 3b) shows a similar pattern, except that now two repeats unfolded first and a single repeat unfolded later. However, the refolding order was the same as in the previous measurement. This observation may reflect some strict refolding order similar to p16 ankyrin repeats, in which carboxy-terminal repeats fold before the amino-terminal repeats²³.

The sixth stretching cycle on the same molecule (Fig. 3c), which was previously stretched to 66 nm, now yielded a maximum extension of 40 nm. This result indicates that two of the three repeats that refolded in the event captured in Fig. 3b (red trace) must also have reattached to the core of the stack and did not unfold in the subsequent pull (see also Supplementary Figs S8 and S9). The fact that the reconstructed stack can sustain forces in excess of 600 pN (Fig. 3c) indicates enormous cohesive forces between neighbouring ankyrin repeats in the stack. This observation also suggests that the (small) force peaks in Fig. 3a, b probably correspond to the

mechanical unfolding of the repeats that detached earlier from the stack and remained separated from it even after refolding. Taken together, these results are consistent with observations of very favourable interfacial interactions between the nearest repeats in the stacks of ankyrin repeats²⁴. The next and subsequent stretching cycles (Fig. 3d, e) revealed three evenly spaced force peaks and two refolding force peaks, indicating that the repeats that had reattached to the stack detached again and were sequentially unfolded. The unfolding patterns shown in Fig. 3 overlap perfectly with the sawtooth pattern of Fig. 2a (Fig. 3f), indicating that the refolding of repeats occurs with high fidelity.

Altogether, we observed a similar refolding behaviour of ankyrin repeats in 75 separate experiments performed on different molecules on different days, using different AFM instruments and cantilevers. In more than 100 separate observations (including different recordings on the same molecule) we recorded major refolding of the stack as shown in Fig. 3c (see also Supplementary Fig. S9d, e). The average refolding force was 32 ± 8 pN (mean $\pm$ s.d., $n = 315$ force peaks; Fig. 3f, inset). Although refolding of proteins under force has been observed previously^{15,25,26}, here we have made direct measurements on single molecules that capture the magnitude of the refolding force of a protein domain.

Ankyrins interact directly with cytoplasmic domains of a variety of membrane transporters and cell adhesion molecules, and couple these proteins to the spectrin-based membrane skeleton²⁷. We propose that the linear elasticity of 24 ankyrin-B repeats shown here might have a role in modulating activity of ankyrin-associated transporters in response to mechanical strain and/or in generating tension in the plane of the membrane bilayer. More generally, the unusually strong tendency of ankyrin repeats to refold may be of particular significance to proteins with four to six ankyrin repeats. Finally, the mechanical properties of ankyrin repeats shown here make them ideal candidates as the building blocks of bio-inspired springy nanostructures and nanomaterials with an inherent ability to self-repair.

METHODS

Purification and characterization of ankyrin-B polypeptides. These are described in Supplementary Information.

Immobilization of ankyrin molecules. Glass coverslips were functionalized with the metal chelate NTA as described in refs 6, 7. NiCl₂ solution (50 μ l, 50 mM) was loaded onto the NTA-functionalized glass for 5 min to achieve

chelation, and the glass was rinsed gently with solution composed of 20 mM phosphate buffer supplemented with 300 mM NaCl; 40 μ l of solution containing heptahistidine-tagged ankyrin molecules at a concentration of 2–10 μ g ml⁻¹ was deposited on a glass coverslip and incubated at 20–25 °C for 20 min.

AFM stretching measurements. All pulling measurements were made with custom-built AFM instruments⁹ equipped with an AFM detector head (Veeco Metrology group) and high-resolution piezoelectric stages (Physik Instrumente) equipped with position sensors (a vertical resolution 0.1 nm). The spring constant of each MLCT-AUHW micro-cantilever (Veeco) was calibrated by using the energy equipartition theorem as described in ref. 28. Molecules were picked up for stretching measurements by gently touching the substrate with the AFM tip, exploiting a nonspecific adsorption of ankyrin to the tip. Force-extension measurements were performed at pulling speeds of between 0.012 and 0.2 nm ms⁻¹, in solution and at room temperature.

Received 28 August; accepted 15 November 2005.

Published online 15 January 2006.

1. Mosavi, L. K., Cammett, T. J., Desrosiers, T. J. & Peng, Z. Y. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**, 1435–1448 (2004).
2. Michaely, P., Tomchick, D. R., Machius, M. & Anderson, R. G. Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J.* **21**, 6387–6396 (2002).
3. Howard, J. & Bechstedt, S. Hypothesis: A helix of ankyrin repeats of the NOMPC-TRP ion channel is the gating spring of mechanoreceptors. *Curr. Biol.* **14**, 224–226 (2004).
4. Corey, D. P. *et al.* TRPA1 is a candidate for the mechanosensitive transduction channel of vertebrate hair cells. *Nature* **432**, 723–730 (2004).
5. Sotomayor, M., Corey, D. P. & Schulten, K. In search of the hair-cell gating spring: Elastic properties of ankyrin and cadherin repeats. *Structure* **13**, 669–682 (2005).
6. Schmid, E. L., Keller, T. A., Dienes, Z. & Vogel, H. Reversible oriented surface immobilization of functional proteins on oxide surfaces. *Anal. Chem.* **69**, 1979–1985 (1997).
7. Schmitt, L., Ludwig, M., Gaub, H. E. & Tampe, R. A metal-chelating microscopy tip as a new toolbox for single-molecule experiments by atomic force microscopy. *Biophys. J.* **78**, 3275–3285 (2000).
8. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
9. Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181–185 (1998).
10. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
11. Fisher, T. E., Marszalek, P. E. & Fernandez, J. M. Stretching single molecules into novel conformations using the atomic force microscope. *Nature Struct. Biol.* **7**, 719–724 (2000).
12. Li, H. B. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
13. Williams, P. M. *et al.* Hidden complexity in the mechanical properties of titin. *Nature* **422**, 446–449 (2003).
14. Oberhauser, A. F., Baddilla-Fernandez, C., Carrion-Vazquez, M. & Fernandez, J. M. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J. Mol. Biol.* **319**, 433–477 (2002).
15. Fernandez, J. M. & Li, H. B. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674–1678 (2004).
16. Rief, M., Pascual, J., Saraste, M. & Gaub, H. E. Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles. *J. Mol. Biol.* **286**, 553–561 (1999).
17. Schwaiger, I., Schleicher, M., Noegel, A. & Rief, M. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO J.* **6**, 46–51 (2005).
18. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of λ -phage DNA. *Science* **265**, 1599–1600 (1994).
19. Corey, D. P. & Sotomayor, M. Tightrope act. *Nature* **428**, 901–903 (2004).
20. Ashmore, J. Channel at the hair's end. *Nature* **432**, 685–686 (2004).
21. Howard, J. & Hudspeth, A. J. Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog's saccular hair cell. *Neuron* **1**, 189–199 (1988).
22. Best, R. B. *et al.* Force mode atomic force microscopy as a tool for protein folding studies. *Anal. Chim. Acta* **479**, 87–105 (2003).
23. Tang, K. S., Fersht, A. R. & Itzhaki, L. S. Sequential unfolding of ankyrin repeats in tumour suppressor p16. *Structure* **11**, 67–73 (2003).
24. Mello, C. C. & Barrick, D. An experimentally determined protein folding energy landscape. *Proc. Natl Acad. Sci. USA* **101**, 14102–14107 (2004).
25. Schwaiger, I., Sattler, C., Hostetter, D. R. & Rief, M. The myosin coiled-coil is a truly elastic protein structure. *Nature Mater.* **1**, 232–235 (2002).
26. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057–2060 (2005).
27. Bennett, V. & Baines, A. J. Spectrin and ankyrin-based pathway: Metazoan inventions for integrating cells into tissues. *Physiol. Rev.* **81**, 1353–1392 (2001).
28. Florin, E. L. *et al.* Sensing specific molecular interactions with the atomic force microscope. *Biosens. Bioelectron.* **10**, 895–901 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Snyder and E. Toone for help with the functionalization of glass coverslips, and B. Sunda and D. Corey for discussions. This work was supported by Duke University funds and grants from the National Science Foundation (to P.E.M.) and the Howard Hughes Medical Institute (to V.B.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.E.M. (pemar@duke.edu) or V.B. (benne012@mc.duke.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Selling ourselves

At most career symposia, panellists describe their career trajectories and offer advice to audience members eager to know how the speakers got to their current positions. But the Nobel Pauling Naturejobs Symposium, held last month in Cambridge, Massachusetts, turned the tables. The panellists offered the usual helpful information. But the audience was also dragged into the act. Grace Wong, the meeting's organizer and chief scientific officer of the Cambridge biotech company Actokine Therapeutics, first encouraged, then cajoled and finally forced each of the 100 or so participants to tell their own career stories.

Wong's method? The 'smart pitch', a technique she has fostered to encourage young scientists to sell themselves — whatever the setting. Wong's *modus operandi* involves a moderator picking an audience member at random and giving them a minute to make an impression. The sometimes reluctant participants are encouraged to tell the audience something memorable about themselves, what they have to offer and what they are seeking — whether it be a job, a collaborator or venture capital.

As the meeting unfolded, each participant got a chance,

welcome or not, to make their pitch. Some of the initial ones were tentative, too long or overly technical. But as the meeting wore on, and became more relaxed, these advertisements grew more succinct, direct and polished.

This was a valuable exercise, because most scientists would admit they could be better at selling themselves. Thinking about how to make a smart pitch before the inevitable introductions can help them to take advantage of opportunities, expected and unexpected, to make connections that could lead to career progress. During the symposium, several young scientists drew interest from panel members looking for new employees, or who knew of colleagues seeking fresh scientific talent. Preparing for these moments is important: you never know when you'll have a microphone — literal or figurative — thrust in front of you.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building,
 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

European Sales Manager:
 Andy Douglas (4975)
Natureevents: Silke Opstrup (4994)

UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
**Germany/Austria/
 The Netherlands:**
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

**European Satellite Office
 Germany/Austria/
 The Netherlands:**
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

MOVERS

Suzanne Fortier, president, Natural Sciences and Engineering Research Council of Canada



2000–05: Vice-principal (academic), Queen's University, Kingston, Ontario, Canada

1995–2000: Vice-principal (research), Queen's University, Kingston, Ontario, Canada

1982–2005: Professor of chemistry and computing, Queen's University, Kingston, Ontario, Canada

Suzanne Fortier's career is built on interactions. As a crystallographer, she found beauty within the inner structures of protein crystals. Now, as leader of Canada's top government funding agency for the sciences, she aims to foster interactions among government, researchers and the public.

Growing up in Quebec at a time when children were educated in religious schools, Fortier's first exposure to science came from a nun who taught chemistry. The nun's infectious enthusiasm spurred Fortier to study science at McGill University in Montreal.

After earning her PhD at McGill, Fortier took a postdoc at the Medical Foundation of Buffalo (now the Hauptman-Woodward Medical Research Institute) in a world-class crystallography lab. She worked under Herbert Hauptman, who pioneered mathematical methods to analyse crystallographic diffraction data. In 1985, he won the Nobel Prize in Chemistry. "He was a fantastic mentor, very supportive and an inspiring person," says Fortier.

In 1982, Fortier was the first woman to be hired by the department of chemistry at Queen's University in Kingston, Ontario. She was later cross-appointed to the department of computing when she recognized a role for artificial intelligence in determining the structure of matter.

Early in her career at Queen's, Fortier joined the Ontario Council on University Affairs, a now defunct university advisory group to the Ontario government. "I was surprised by the richness of the interactions," she says. "I had had no interest in administration, but I learned how exciting teamwork and policy-making were."

That experience encouraged her to make the jump from research to administration at Queen's in the mid-1990s, where she became vice-principal (research) and later vice-principal (academic). She missed research and teaching, but found that contributing to the institution as a whole provided a sense of accomplishment. "As administrators, we are there to serve our community," she says. "The success of others is what has to motivate you."

After 11 years as vice-principal, Fortier moved to Ottawa in January to head the Natural Sciences and Engineering Research Council. Her aim is to encourage the success of researchers while also enhancing communication between them and the Canadian public. "It's a goal of all our universities to be players on the world stage," she says, "so how can we all work together to achieve that?" ■

Hannah Hoag

SCIENTISTS & SOCIETIES

Actors as teachers

Scientists often have exceptional technical skills but may have received little training in communication and management. To address this gap, a new training programme called LabAct was introduced last year by the Laboratory Management Institute (LMI) at the University of California, Davis. It uses professional actors to help young scientists rehearse behaviour that can be effective in dealing with difficult communication, management and ethical situations they may face in their laboratories. We believe that this training will enhance overall productivity, quality and job satisfaction.

The actors receive specialized training in techniques for drawing out and engaging participants and helping them gain confidence. They use their expertise to help participants identify real workplace issues and practise new behaviours to resolve them.

Last year, 22 scientists, mainly postdocs, attended a two-day LMI training session where speakers gave talks on topics such as leadership, ethics and project management. LabAct was one part of the workshop. Participants anonymously submitted issues they wanted to see enacted, such as conflicts over authorship, access to shared research equipment, and work habits. The actors improvised scenes that illustrated the issues, often

in an egregious or humorous manner. After a group discussion, the actors re-enacted the scene incorporating participant-recommended behaviour. This allowed the participants to experiment with different behaviours, through the actors, until they found those that resonated best with them. As participants felt more comfortable, some joined the actors to participate in the scene, gaining practice in effective and authentic communication.

What distinguishes LabAct training from online instruction or discussion of case studies is that LabAct gives participants hands-on practice in developing new communication and problem-solving skills in a safe environment before implementing them in the workplace.

Originally conceived for postdoctoral scholars and new investigators, LabAct training can be used for many topics and audiences. We are now developing LabAct training to attract young students and members of under-represented groups to careers in science, and have also been approached by industry and government to provide this training. ■

Jade McCutcheon is assistant professor of theatre, and John Galland is director of the Laboratory Management Institute, the University of California, Davis.
♦ www.research.ucdavis.edu/LMI

GRADUATE JOURNAL

Tropical PhD

Growing up in Sweden, I realized that I did not like cold weather. My mother used to read me a book about a penguin that had the same problem. One day he decided he had had enough, so he set sail towards the tropical Pacific in a bathtub. I did the same thing, although I did not use a bathtub. I thought life would be much more pleasant in Hawaii. I was right.

To secure my status as a subtropical resident for several years, I decided to pursue a PhD in oceanography at the University of Hawaii. Science, combined with some fun under the sun, seemed like an excellent idea. To limit the number of days spent in cold weather, I was drawn to the study of coral reefs; more precisely, the study of the effects of rising atmospheric carbon dioxide on this ecosystem. Sunny skies, warm waters, beautiful scenery and the most costly experiment mankind has ever undertaken: burn all fossil fuels and just sit back and watch. Could life get any better than this?

But recently it struck me that the day will soon come when I might see an end to both an initially endless PhD and the coral reefs as we know them, leaving me the question of what to do next. I have great hopes of presenting you with the answer to that before the end of the year. Although life is still very pleasant in Hawaii, I am getting the bathtub ready. The quest for a new destination has begun. ■

Andreas Andersson is a final-year PhD student in oceanography at the University of Hawaii.

The punishment fits the crime

Everything's going to be all right.

David Berreby

Day 1: Come in, Tommy. Sit down. Don't be afraid. I have good news! Does your head hurt right now? Has it hurt any time since the doctor saw you? No, don't be scared. I have something wonderful to tell you: you're leaving us soon. Very soon. Thanks to the doctor, you're going to be smart!

Yes indeed, as smart as I am. Actually, you'll probably be smarter. (You always were, years ago, and you never let me forget it.)

Angry? Not at all. You have nothing to be afraid of! You'll see.

I think that's enough for now. I'll see you next week.

Day 8: Good morning, Tom. How's your head? Good. We don't expect those pains to last much longer. But now that you've matured a bit, I can be frank with you: we don't know much about what's happening to you, because you're one of the first people to be released from the programme. In fact, you're number one!

By now, I assume you understand that you're here because you did something wrong. No, no, sit down. Really, you must try to stay calm. It was years ago, and, as you've learned, it's all squared away now. Your debt is paid.

What? Well, you could say that, but I must say that's not a very positive way to look at things. You couldn't handle life on the outside, until a few days ago. Don't you remember? You thought and acted like a very small child, until we began the treatments.

I know! You couldn't help it, of course. That was your sentence: three year's severe impairment. Which I, on behalf of the duly constituted authorities, legal and medical, am pleased to say is ending.

No, I think that's enough for now.

Day 15: Well. That was quite an outburst, but I will try to be patient. Your position is unusual. One might say unique.

What do you want to know, then?

Yes, your sentence was entirely legal and approved only after expert review. No 'poisons' were used. What an idea! You were given the best treatments we have — the vectors are absolutely safe, and our techniques here for somatic-genetic engineering are the best in the world. As you know. Or knew.

Of course it's not funny. What makes me



smile is pride — pride at the hard work of so many brilliant scientists, and the wisdom of our correctional reforms. Times were so different then. We thought the country was falling to pieces, as you may remember.

You don't? Ah, well.

The prisons were ever more crowded and expensive, even as genomic medicine was ever more successful and economical. It was only a matter of time before someone saw that techniques for gene therapy could also deliver gene punishment.

I was simply in the right place at the right time. I make no claim to brilliance or foresight.

And now, just a few years later, genetic correction is the law. In your case, you now produce a toxin that kills new-forming brain cells. Naturally, this has some effect on your mood, and your memory. And the expression of genes in your frontal cortex has been, shall we say, gently reorganized, so that your level of self-control has been

approximately that of a normal three-year old. Of course, you had to be cared for, and so you came to live here with us. This is not a prison, but a facility for research and compassionate care. A facility I am proud to direct, I might add.

Oh, my. Your self-control doesn't seem to be coming back so well. That's all for today. Take him away.

Day 23: I must say my feelings are mixed. I am glad to see that you have recovered your abilities so completely: abilities which were once the envy of your colleagues, among whom, as you now recall, I number my humble self. But these anti-social remarks are very troubling.

Once again, I must remind you that you were sentenced after a lawful and constitutionally sound trial. And as sorry as you seem to feel for yourself now, I think, frankly, that you had it easy. Although the state no longer maintains a prison system, you have been provided with care and supervision. Most convicts are not. Do you know what the sentence for armed robbery in this state is today? Five years of lower-body paralysis, and good luck to you. Tax evasion? Two to five years' severe autism. Many families find it to be quite a burden. No, I'm sorry, I don't have to listen to such nonsense. Cruel and unusual, my foot! The High Court ruled against you: yes, against you. A leading geneticist campaigning against genetic punishment, confusing the moral and legal issues, making the people of this nation think that scientists weren't united in support of essential reforms. After passage of the Cellular Corrections Law of 2019, your activities were, in fact, criminal.

Ah, well. I'm afraid your disagreement doesn't matter much to me, Thomas. And I will have to ask you to set it aside.

No, that's not exactly right. It's not up to you. There is the little matter of probation. Your treatments here must continue: you will need a fresh injection every month for the rest of your life, and if you do not meet the conditions of your parole, such an injection will not be forthcoming.

From me. As director here, and as your friendly colleague of former days, I have taken it upon myself to monitor your progress.

I trust you will comport yourself with tact, then? Good. See you next month. ■

David Berreby is a science writer and the author of *Us and Them*, a non-fiction book about group identity. He lives in New York City.

JACEY